AD-A031 752 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/2
ON SOME NEW BIVARIATE NEGATIVE BINOMIAL AND GAMMA DISTRIBUTIONS--ETC(U)
FEB 76 C R MITCHELL
UNCLASSIFIED AFIT-CI-7T-8 NL

1 OF 2
AD
A031752

CI-77-8

1

ON SOME NEW BIVARIATE NEGATIVE BINOMIAL AND GAMMA DISTRIBUTIONS WITH APPLICATIONS TO QUEUES, INVENTORIES AND MAINTENANCE

by

Charles R. Mitchell

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: Operations Research and Statistics

Approved by the
Examining Committee:

Albert S. Paulson

Albert S. Paulson, Thesis Adviser

George E. Manners, Jr., Member

Pasquale Sullo, Member

John W. Wilkinson, Member

Rensselaer Polytechnic Institute
Troy, New York

February 1976
(For graduation May 1976)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
CI 7T-8

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
6
On Some New Bivariate Negative Binomial and Gamma Distributions With Applications to Queues, Inventories and Maintenance.

5. TYPE OF REPORT & PERIOD COVERED
PhD Thesis

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
10
CHARLES R. MITCHELL
MAJOR, USAF

8. CONTRACT OR GRANT NUMBER(s)
9
Doctoral thesis,

9. PERFORMING ORGANIZATION NAME AND ADDRESS
AFIT Student at Rensselaer Polytechnic Institute
Troy, New York

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
AFIT/CI
Wright-Patterson AFB OH 45433

12. REPORT DATE
11
Feb[redacted]76

13. NUMBER OF PAGES
144 pages

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)
12 156p.

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for Public Release; Distribution Unlimited
14 AFIT-CI-7T-8

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES
APPROVED FOR PUBLIC RELEASE AFR 190-17.
Jerral F. Guess
JERRAL F. GUESS, Captain, USAF
Director of Information, AFIT

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)
Attached

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 012 200

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)
LB

SECURITY CLASSIFICATION OF THIS PAGE(*When Data Entered*)

SECURITY CLASSIFICATION OF THIS PAGE(*When Data Entered*)

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

ACKNOWLEDGEMENT

I gratefully acknowledge the guidance and encouragement provided by Professor Albert S. Paulson throughout this research and thank him for making it such a rewarding experience. Also, I appreciate the encouragement received during my graduate studies from the other members of my Doctoral committee, Professors George E. Manners, Jr., Pasquale Sullo, and John W. Wilkinson. Special thanks go to my wife, Sylvia, and children, Charles, Leana and Bruce, for their patience and understanding.

I am grateful to the United States Air Force and particularly to the Department of Mathematics, United States Air Force Academy for sponsoring this program.

ABSTRACT

Some new theory and applications of certain bivariate non-normal distributions are presented. In particular, new bivariate negative binomial and gamma distributions are discussed and an existing bivariate exponential distribution is applied to single stage and tandem queueing systems (both single server) which have particular kinds of correlated structure.

One new bivariate negative binomial distribution is derived by convoluting an existing bivariate geometric distribution; the probability function has six parameters and admits of positive or negative correlations and linear or non-linear regressions. Given are the moments to order two and for special cases, the regression function, a recursive formula for the probabilities, a method of moments parameter estimation technique, the likelihood equations, the differential-difference equations and for maximum likelihood parameter estimates, a necessary relationship for the parameters. Certain results are extended to a dual bivariate gamma distribution. Another bivariate negative binomial distribution, which has four parameters, results by reducing a particular trivariate negative binomial distribution with independent marginals; only positive correlations and linear regressions are possible here. Both bivariate negative binomial distributions are

fitted to data and their special features illustrated.

Applications of bivariate distributions to certain aircraft logistical problems are investigated. Primarily, a bivariate negative binomial distribution is fitted to spare parts demand data in two periods and to monthly abort data on either side of a large scale maintenance event and it is shown how the associated sample distributions can be useful in parts inventory control and in investigating the effect of maintenance on an aircraft's performance.

A bivariate exponential distribution is applied to tandem queues to study the effect of correlated exponential service times and to single stage queues to study the effect of correlation between a customer's service time and the interarrival interval separating himself and his predecessor. Arrivals to both systems are according to a Poisson process. Simulation is used to show that the mean waiting time is quite sensitive to departures from the traditional assumptions of mutually independent service times for tandem queues and independence of service times and interarrival intervals for single stage queues, especially at higher utilizations. For the cases of infinite interstage storage between two-stage tandem queues and infinite storage before a single stage queue, system performance is increased by positive correlation and impaired by negative correlation. For two-stage queues this change is reversed for zero interstage storage and depends on the value of

the utilization rate for the case where interstage storage equals unity. By using spectral analysis techniques and a nonparametric test applied to sample power spectra associated with certain simulated waiting times the effects are shown to be statistically significant. For correlation equal unity and infinite interstage storage results are given for two through twenty-five stages in series.

PART 1

INTRODUCTION

The objectives of this research are (1) to develop two new bivariate negative binomial probability functions, (2) to derive, where applicable, corresponding properties for a dual bivariate gamma distribution, and (3) to show that bivariate approaches to data analysis and mathematical modeling can provide, in some cases, more meaningful and representative results than traditional approaches. The first bivariate negative binomial (bnb) distribution is derived, via convolution, from an existing bivariate geometric distribution and the second one is developed, via reduction, from a certain independent trivariate negative binomial distribution. Certain properties related to infinite divisibility and parameter estimation are shown to be directly applicable to an existing bivariate gamma distribution.

In analyzing bivariate data from self-pairing type studies usually the data are transformed to obtain a univariate random variable; sometimes this technique is acceptable particularly if the data are approximately normal but often a bivariate random variable, say, representing count data, cannot be adequately treated in this way and so an alternate approach is deemed necessary. Most often in these cases a first step is to find a reasonable bivariate probability function to represent the data and for this purpose several probability functions

are usually compared. That these new bnb distributions should be useful in practical applications is suggested by showing, for two bivariate samples from the literature, how certain properties of these distributions better represent the data. Additionally, we introduce some new bivariate data sets related to aircraft operations and maintenance and show how bivariate approaches can be useful in analyzing certain problems dealing with these data.

Another new bivariate approach is related to correlated queueing systems. For instance, single server tandem queues traditionally have been modeled by assuming that customer service times at the individual servers are independent; sometimes this is a reasonable assumption but in certain important applications, for example, production lines, such is often not the case and a more realistic model is desired. We show for two servers in series the effect of correlated exponential service times by assuming that a customer's service times at the two servers are given by a bivariate exponential distribution. Also we provide a similar analysis for single server, single stage queues with correlated interarrival and service processes.

Part 2 describes the new bnb distributions with certain associated results for a bivariate gamma distribution and also shows how the bnb distributions fit some empirical data. We introduce in Part 3 some new discrete bivariate data sets

related to aircraft operations and maintenance and show how bnb distributions can be useful in analyzing certain problems dealing with these data. Also we describe how bivariate gamma distributions could be useful in similar problems associated with continuous data. Part 4 shows the results on correlated queueing systems. More introductory and historical remarks are included in these parts.

PART 2

ON BIVARIATE NEGATIVE BINOMIAL (AND GAMMA) DISTRIBUTIONS

2.1 Introduction and Historical Review

In this part we develop and fit to data two new bnb distributions and show how certain properties relate to an existing bivariate gamma distribution.

That a bnb should be important in statistical theory and applications is suggested by the wide acceptance of the univariate negative binomial distribution as a reasonable model for a broad range of problems representing univariate discrete random variables (see Boswell and Patil (1970) for a discussion of fifteen stochastic models which give rise to the univariate negative binomial distribution). It is reasonable to suspect that a bnb distribution would be useful in describing bivariate random variables for which correlation exist between the members of the bivariate pair and the marginals are negative binomial. A few bnb distributions have been presented in the literature; next we show the particular forms of the univariate negative binomial to be used herein and then review these bnb distributions and some of their applications.

The univariate negative binomial with parameters $\nu>0$ and $\theta>0$ is defined (Johnson and Kotz (1969)) as the distribution of a random variable (r.v.) X for which

$$\Pr[X=x] = \frac{(x+\nu-1)!}{x!(\nu-1)!}\left(\frac{1}{1+\theta}\right)^{\nu}\left(\frac{\theta}{1+\theta}\right)^{x},\quad x=0,1,2,\ldots \tag{2.1}$$

and the characteristic function is

$$\phi(t) = E[e^{itX}] = [1+\theta(1-e^{it})]^{-\nu}. \tag{2.2}$$

The mean and variance of X are $\nu\theta$ and $\nu\theta(1+\theta)$, respectively. Thus, it is characteristic of the negative binomial distribution that the variance is greater than the mean. A method of moments parameter estimation technique is described in Johnson and Kotz. For $\nu=1$ we have the geometric distribution.

Another common representation is to let $\theta=p/(1-p)$, or equivalently, $p=\theta/(1+\theta)$, in (2.1) and so

$$\Pr[X=x] = \frac{(x+\nu-1)!}{x!(\nu-1)!}(1-p)^{\nu}\, p^{x},\quad x=0,1,2,\ldots \tag{2.3}$$

and

$$\phi(t) = [1+\frac{p}{1-p}(1-e^{it})]^{-\nu} = [\frac{1-pe^{it}}{1-p}]^{-\nu}. \tag{2.4}$$

This latter representation is referred to as a negative binomial distribution with parameters ν and p. We use both forms throughout.

The negative binomial distribution can be viewed as a compound distribution (Johnson and Kotz). In fact, a mixture of Poisson distributions such that the expected values, Λ, of the Poisson distributions vary according to a Type III (gamma) distribution with probability density function

$$f(\lambda) = (\tfrac{1}{\theta})^{\nu} \frac{1}{\Gamma(\nu)} \lambda^{\nu-1} e^{-\lambda/\theta}, \ \lambda>0;\ \nu>0,\ \theta>0 \qquad (2.5)$$

leads to 2.1. A multivariate negative binomial distribution was constructed in an analogous way by Bates and Neyman (1952). We describe the bivariate case. Suppose we consider the joint distribution of the two independent random variables X and Y where X is distributed as a Poisson r.v. with expected value Λ and Y is also a Poisson r.v. but its expected value is $\alpha\Lambda$, $\alpha>0$ and constant. If we assume that Λ is distributed according to the gamma distribution in (2.5), then the marginal distribution of X and Y is

$$\Pr[X=x, Y=y] = \frac{(x+y+\nu-1)!}{x!y!(\nu-1)!} (1-p-q)^{\nu} p^x q^y, \ x,y=0,1,2,\ldots \qquad (2.6)$$

where $p=\theta/[1+(\alpha+1)\theta]$, $q=\alpha p$, $\nu>0$, $0 < p < 1$, $0 < q < 1$, and $0 < p+q < 1$. The characteristic function is

$$\phi(t_1,t_2) = E[e^{it_1X+it_2Y}] = [\frac{1-pe^{it_1}-qe^{it_2}}{1-p-q}]^{-\nu}, \qquad (2.7)$$

or in terms of θ,

$$\phi(t_1,t_2) = [1+\theta(1-e^{it_1}) + \alpha\theta(1-e^{it_2})]^{-\nu}. \qquad (2.8)$$

We have that the mean vector is

$$\underline{\mu} = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} \nu\theta \\ \nu\alpha\theta \end{bmatrix} \qquad (2.9)$$

and the covariance matrix is

$$\Sigma = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} \nu\theta(1+\theta) & \nu\alpha\theta^2 \\ \nu\alpha\theta^2 & \nu\alpha\theta(1+\alpha\theta) \end{bmatrix} . \tag{2.10}$$

From the characteristic function it is clear that the marginals are again negative binomial. The conditional distribution can be shown to be

$$\Pr[Y|x] = \frac{(x+y+\nu-1)!}{y!(x+\nu-1)!}(1-q)^{x+\nu} q^y, \; y=0,1,2,... \tag{2.11}$$

or a negative binomial with parameters $x+\nu$ and q. Therefore, the conditional mean or regression function is

$$E[Y|x] = q(\nu+x)/(1-q) \tag{2.12}$$

and we note that the form is linear. Note also that this probability function admits of positive correlations only.

Besides Bates and Neyman in 1952 others have studied the above bnb distribution (Mardia (1970) gives an historical review). Guldberg introduced this distribution in 1934, Lundberg first used it in 1940 as a model for accident proneness and Arbous and Kerrich (1951) expanded Guldberg's work and fitted the distribution to bivariate data related to accidents in industrial settings. In addition to their theoretical contributions Bates and Neyman fitted the distribution to numerous bivariate data sets related to diseases in industrial workers. In 1954 Arbous and Sichel fitted it to shift-worker absenteeism data for adjacent time periods. Youngs, Geisler, and Brown in

1955 studied the conditional distribution of this bivariate r.v. and showed how it could be used for the prediction of demand for aircraft spare parts. In 1961 Edwards and Gurland generalized (2.6) and compared the fits obtained from the two distributions. The regression function for their distribution is linear also. Subrahmaniam (1966) and Subrahmaniam and Subrahmaniam (1973) also studied this latter bnb distribution.

For purposes of comparison we choose to call the bnb distribution in (2.6) the Guldberg-Bates-Neyman model with parameters α, θ and ν and to designate it G-B-N(α,θ,ν).

Certain data sets do not exhibit empirical regressions which are linear nor do some data sets show positive correlation and so it is natural, for these cases, to work with a bivariate probability function which allows for nonlinear regressions or negative correlations or both. The classical Bates and Neyman paper exhibited empirical data best fitted by regression curves which were obviously nonlinear, and, consequently, their results were not entirely satisfactory. Furthermore, we show a new bivariate data set related to aircraft flight aborts which has a negative sample correlation coefficient.

In the next section we discuss a new bnb distribution which admits of nonlinear regressions and negative correlations and derive several of its properties.

2.2 A New Bivariate Negative Binomial (and Gamma) Distribution, Via Convolution, and Some Properties

The purpose of this section is to show how a new bnb distribution may be obtained, by the process of convolution, from a certain bivariate geometric distribution. A number of properties such as the moments to order two, the regression function, a recursive formula for the cell probabilities, and the likelihood equations are obtained for certain special cases.

Paulson and Uppuluri (1972) showed that the bivariate r.v. (X,Y), where each element in the pair is defined on the non-negative integers, has a bivariate geometric distribution if its characteristic function, $\phi(t_1,t_2)$, satisfies the characteristic-functional equation

$$\phi(T) = \psi(T)\ E[\phi(TV)] \tag{2.13}$$

where

$$T = (t_1,t_2),\ \psi(T) = \psi_1(t_1,0)\ \psi_2(0,t_2),$$

$$\psi_1(t_1,0) = [1+\frac{p}{1-p}\ (1-e^{it_1})]^{-1},$$

$$\psi_2(0,t_2) = [1+\frac{q}{1-q}\ (1-e^{it_2})]^{-1},$$

and V is a 2x2 matrix-valued r.v. having values in the set $\{ \begin{bmatrix}0&0\\0&0\end{bmatrix}, \begin{bmatrix}1&0\\0&0\end{bmatrix}, \begin{bmatrix}0&0\\0&1\end{bmatrix}, \begin{bmatrix}1&0\\0&1\end{bmatrix} \}$ with probabilities a,b,c and d, respectively. Here $0 < p < 1$, $0 < q < 1$, $a+b+c+d=1$, $b+d<1$ and $c+d<1$. Thus the characteristic functional equation can be rewritten as

$$\phi(T) = \psi_1(t_1,0)\ \psi_2(0,t_2)[a+b\phi(t_1,0) + c\phi(0,t_2) + d\phi(T)] \quad (2.14)$$

and it is easy to show that

$$\phi(t_1,0) = \frac{(a+c)\psi_1}{1-(b+d)\psi_1} = [1+\theta_1(1-e^{it_1})]^{-1},$$
$$\phi(0,t_2) = \frac{(a+b)\psi_2}{1-(c+d)\psi_2} = [1+\theta_2(1-e^{it_2})]^{-1} \quad (2.15)$$

where

$$\theta_1 = p/[(1-p)(a+c)],\ \theta_2 = q/[(1-q)(a+b)] \quad (2.16)$$

and the arguments of $\psi_1(t_1,0)$ and $\psi_2(0,t_2)$ have been suppressed (and will be in the sequel). Comparing (2.15) and (2.2) we see that the marginals are geometric.

In (2.13) the characteristic function $\psi(T) = \psi_1(t_1,0)\cdot\psi_2(0,t_2)$ corresponds to a bivariate geometric distribution with independent geometric marginals and as Block (1975) points out, (2.13) gives the characteristic-functional equation of the bivariate random variable

$$(X,Y) = \left(\sum_{i=1}^{N_1} X_i,\ \sum_{i=1}^{N_2} Y_i\right). \quad (2.17)$$

The pair (N_1,N_2) is a certain bivariate geometric distribution and is independent of (X_i,Y_i), i=1,2,3,..., which are independent and identically distributed (i.i.d.) random variables with characteristic function $\psi(T)$. Thus (2.13) corresponds to a special kind of bivariate geometric compounding of the distribution with characteristic function $\psi(T)$. It is also clear that

if $\psi(T)$ has geometric marginals, then $\phi(T)$ will have geometric marginals. This is even more apparent from (2.17) since univariate random geometric sums of i.i.d. geometric random variables are geometric.

Paulson (1973) has shown that (2.13) also characterizes a bivariate exponential distribution where $\psi(T)$ is the characteristic function of independent exponential r.v.'s. In addition he forms, in a way to be described here for a bnb distribution, a bivariate gamma distribution. Certain properties to be derived for this bnb distribution will be extended to his bivariate gamma distribution.

Paulson and Uppuluri obtained the moments of the distribution of (X,Y) in (2.13) to order two and showed that the correlation coefficient has values in the interval $-0.25 \leq \rho < 1$. They also presented recursive formulae for determining the probability function.

Clark (1972) obtained a closed form representation for the bivariate geometric distribution characterized by (2.13) for the special case b=c=0 and extended it to a bnb distribution. Next we summarize that development. (By defining a multivariate analogue of the characteristic-functional equation (2.13), Clark also constructed a multivariate geometric distribution and extended it to a multivariate negative binomial distribution; Appendix A shows the unpublished derivation.)

We consider the case b=c=0 in (2.14) and so

$$\phi(T) = \psi_1\psi_2(a+d\phi(T)). \tag{2.18}$$

Solving for $\phi(T)$ leads to

$$\phi(T) = a\psi_1\psi_2[1-d\psi_1\psi_2]^{-1} \tag{2.19}$$

and upon expansion

$$\phi(T) = a\psi_1\psi_2[1+d\psi_1\psi_2 + (d\psi_1\psi_2)^2 + \ldots] . \tag{2.20}$$

The inverse transform of $\phi(T)$, that is, the probability function, say, $g_1(x,y)$, may be obtained termwise from (2.20) since the resultant series converges uniformly and absolutely for all $x,y=0,1,2,\ldots$, (Titchmarsh(1964)); we have

$$g_1(x,y) = a(1-p)p^x(1-q)q^y \sum_{j=0}^{\infty} \binom{x+j}{x}\binom{y+j}{y}[d(1-p)(1-q)]^j \tag{2.21}$$

where $x,y=0,1,2,\ldots$. Expanding the combinatorial terms gives

$$g_1(x,y) = a(1-p)p^x(1-q)q^y \; F(x+1,y+1;\ 1;\ d(1-p)(1-q)) \tag{2.22}$$

where $F(a,b;c;z)$ is the Gaussian hypergeometric series given by

$$F(a,b;c;z) = 1 + \sum_{j=1}^{\infty} \frac{(a)_j(b)_j}{(c)_j}\frac{z^j}{j!} \tag{2.23}$$

and the term $(n)_j$ is defined by

$$(n)_j = \frac{\Gamma(n+j)}{\Gamma(n)} = n(n+1)(n+2)\ldots(n+j-1). \tag{2.24}$$

Taking the ν-fold convolution of $g_1(x,y)$ with itself yields a bnb distribution which is denoted by $g_\nu(x,y)$. The characteristic function of $g_\nu(x,y)$ is

$$\phi_\nu(T) = [\phi(T)]^\nu = (a\psi_1\psi_2)^\nu[1-d\psi_1\psi_2]^{-\nu}$$

$$= (a\psi_1\psi_2)^\nu[1+\nu(d\psi_1\psi_2) + \frac{\nu(\nu+1)}{2!}(d\psi_1\psi_2)^2 + \ldots] \qquad (2.25)$$

and in the same manner as before we obtain

$$g_\nu(x,y) = a^\nu h_1(x)h_2(y)\ F(x+\nu,y+\nu;\nu;z) \qquad (2.26)$$

where

$$h_1(x) = \binom{\nu+x-1}{x}(1-p)^\nu p^x,$$

$$h_2(y) = \binom{\nu+y-1}{y}(1-q)^\nu q^y, \qquad (2.27)$$

$$z = d(1-p)(1-q)$$

and $x,y=0,1,2,\ldots$. Of course, $\nu=1$ leads to (2.22). It can be shown that

$$\phi_\nu(t_1,0) = [1+\theta_1(1-e^{it_1})]^{-\nu}$$

and

$$\phi_\nu(0,t_2) = [1+\theta_2(1-e^{it_2})]^{-\nu} \qquad (2.28)$$

where θ_1 and θ_2 are defined as in (2.16) and so the marginals of (X,Y) are negative binomial.

Clark obtained the moments to order two of this bnb distribution; the remainder of his work was limited to numerical investigations of the bivariate geometric distribution characterized by (2.14) with $b \neq 0$ and $c \neq 0$, and he showed figures of the probability surface and the regression function as it depends upon ρ, the correlation coefficient. The probability surface was computed by using the recursive formulae given by Paulson and Uppuluri and then the regression function was computed by using the definition of a conditional mean, that is, $E[Y|x] = \sum y \Pr(y|x)$, where the summation is taken over all non-negative y. His numerical results showed that the probability function admits of nonlinear regression functions with either positive or negative correlations.

Next we generalize Clark's work for the cases b and c not equal zero and show, among other things, an analytical derivation of the regression function for some special cases.

We construct in a way parallel to Paulson's (1973) derivation for a bivariate gamma distribution a bnb distribution. From (2.14)

$$\phi(T) = \psi_1\psi_2[a + b\phi_1 + c\phi_2 + d\phi(T)] \tag{2.29}$$

or

$$\phi(T) = \frac{\psi_1\psi_2}{(1-d\psi_1\psi_2)}[a + b\phi_1 + c\phi_2] \tag{2.30}$$

and convolving as in (2.25) yields

$$\phi_\nu(T) = \left(\frac{\psi_1\psi_2}{1-d\psi_1\psi_2}\right)^\nu [a + b\phi_1 + c\phi_2]^\nu . \qquad (2.31)$$

From (2.29) on we write ϕ_1 and ϕ_2 for $\phi(t_1,0)$ and $\phi(0,t_2)$, respectively. Recalling from (2.13) that ψ_1 and ψ_2 are functions of p and q respectively, we choose to designate the particular bnb distribution which results here as the <u>BNB(a,b,c,p,q,ν)</u> distribution and to label the probability function as <u>$f_\nu(x,y)$</u>. It is relatively easy to show from (2.31) that the marginals are negative binomial. The following theorems and results are now presented.

<u>Theorem 1</u>: The inverse transform of $\phi_\nu(T)$ defined by (2.31) with a,b,c,d as probabilities and a+b+c+d=1, b+d<1, c+d<1 and ν integer is

$$f_\nu(x,y) = \sum_{\alpha,\beta,\gamma} \frac{\nu!}{\alpha!\beta!\gamma!} a^{\alpha-\nu} b^\beta c^\gamma \Phi_{1.\beta} \overset{x}{*} g_\nu(x,y) \overset{y}{*} \Phi_{2.\gamma} \qquad (2.32)$$

where $\sum$ runs over all $\alpha,\beta,\gamma \geq 0$ such that $\alpha + \beta + \gamma = \nu$, x,y=0,1,2,..., $g_\nu(x,y)$ is the probability function for the BNB(a,0,0,p,q,ν) of equation (2.26) and

$$\Phi_{1.\beta} = \left(\frac{1}{1+\theta_1}\right)^\beta \binom{x+\beta-1}{x} \left(\frac{\theta_1}{1+\theta_1}\right)^x ,$$

$$\Phi_{2.\gamma} = \left(\frac{1}{1+\theta_2}\right)^\gamma \binom{y+\gamma-1}{y} \left(\frac{\theta_2}{1+\theta_2}\right)^y . \qquad (2.33)$$

The operator $\overset{x}{*}$ for convolution over x is defined for two functions $h_1(x,y)$ and $h_2(x,y)$ by

$$h_1 \overset{x}{*} h_2 = \sum_{\xi=0}^{x} h_1(\xi,y)h_2(x-\xi,y) . \qquad (2.34)$$

The operator $\overset{y}{*}$ is defined similarly.

Proof: From (2.31) and using the multinomial expansion gives

$$\phi_\nu(T) = \left(\frac{\psi_1\psi_2}{1-d\psi_1\psi_2}\right)^\nu \sum_{\alpha,\beta,\gamma} \frac{\nu!}{\alpha!\beta!\gamma!} a^\alpha (b\phi_1)^\beta (c\phi_2)^\gamma \quad (2.35)$$

where the $\sum$ is over all $\alpha,\beta,\gamma \geq 0$ such that $\alpha + \beta + \gamma = \nu$. Recall from (2.15) that $\phi_1 = [1+\theta_1(1-e^{it_1})]^{-1}$ and $\phi_2 = [1+\theta_2(1-e^{it_2})]^{-1}$ and changing to z-transforms by letting $e^{it_1} = r^{-1}$ and $e^{it_2} = s^{-1}$ we can write

$$\phi_{1r} = \frac{1}{1+\theta_1}\left[\frac{r}{r-\frac{\theta_1}{1+\theta_1}}\right] \text{ and } \phi_{2s} = \frac{1}{1+\theta_2}\left[\frac{s}{s-\frac{\theta_2}{1+\theta_2}}\right]$$

where ϕ_{1r} and ϕ_{2s} are the corresponding transformations. Taking the inverse z-transform, $\mathcal{Z}^{-1}$, of ϕ_{1r}^β and ϕ_{2s}^γ we get (from the z-transform pair number 24, Jury(1964))

$$\mathcal{Z}^{-1}(\phi_{1r}^\beta) = \Phi_{1.\beta} = \left(\frac{1}{1+\theta_1}\right)^\beta \binom{x+\beta-1}{x}\left(\frac{\theta_1}{1+\theta_1}\right)^x, \ \beta \geq 1, \ \Phi_{1.0}=1,$$

and

$$\mathcal{Z}^{-1}(\phi_{2s}^\gamma) = \Phi_{2.\gamma} = \left(\frac{1}{1+\theta_2}\right)^\gamma \binom{y+\gamma-1}{y}\left(\frac{\theta_2}{1+\theta_2}\right)^y, \ \gamma \geq 1, \ \Phi_{2.0}=1.$$

Except for the constant a^ν, the inverse transform of $\left(\frac{\psi_1\psi_2}{1-d\psi_1\psi_2}\right)^\nu$ is $g_\nu(x,y)$ in equation (2.26) so (2.35) may readily be inverted (after changing to z-transforms throughout) by first holding s, say, constant and inverting with respect to r and then completing

the inversion by inverting with respect to s. Thus we arrive at (2.32). This equation is analogous to Paulson's (1973) result for a bivariate gamma distribution (a^{α} in equation (26) of that paper should be replaced with $a^{\alpha-\nu}$). Directly from Theorem 1 we have

Corollary 1: For $\nu=1$ and $a\neq 0$, the probability function for the BNB(a,b,c,p,q,1) distribution is

$$f_1(x,y) = g_1(x,y) + \frac{b}{a}\left(\frac{1}{1+\theta_1}\right)\left(\frac{\theta_1}{1+\theta_1}\right)^x \overset{x}{*} g_1(x,y)$$
$$+ \frac{c}{a}\left(\frac{1}{1+\theta_2}\right)\left(\frac{\theta_2}{1+\theta_2}\right)^y \overset{y}{*} g_1(x,y) \qquad (2.36)$$

where $g_1(x,y)$ is the probability function for the BNB(a,0,0,p,q,1) of (2.21). This is a closed form representation of Paulson and Uppuluri's bivariate geometric distribution.

Obviously, the utility of $f_\nu(x,y)$ in (2.32) is limited by the integral requirement for ν and so we seek a representation for ν real valued. Except for the case c=0 our attempts to derive a general representation have been unsuccessful. Next we give the results for c=0 (or for b=0 by symmetry). Also, we show how these results are directly applicable to Paulson's bivariate gamma distribution.

Theorem 2: For the BNB(a,b,0,p,q,ν) distribution with $\nu>0$,

$$f_\nu(x,y) = g_\nu(x,y) + g_\nu(x,y) \overset{x}{*} \sum_{k=1}^{\infty} h(b,k)(1-p)^k\binom{x+k-1}{x}p^x \qquad (2.37)$$

where $g_\nu(x,y)$ is the BNB$(a,0,0,p,q,\nu)$ of (2.26) and

$$h(b,k) = [-(b+d)]^k \sum_{n=0}^{k} \binom{-\nu}{k-n}\binom{\nu}{n}\left(\frac{d}{b+d}\right)^n. \tag{2.38}$$

Proof: In (2.31) we write $\phi_\nu(c=0)$ for the characteristic function when c=0 and by using (2.15) it follows that

$$\phi_\nu(c=0) = \left(\frac{a\psi_1\psi_2}{1-d\psi_1\psi_2}\right)^\nu \left[1+\frac{b\psi_1}{1-(b+d)\psi_1}\right]^\nu. \tag{2.39}$$

The leading term in this expression is the characteristic function for the case b=c=0 and for it we write $\phi_\nu(b=c=0)$. Rewriting the second term and expanding in an infinite series gives

$$\begin{aligned}\phi_\nu(c=0) &= \phi_\nu(b=c=0)\left\{\frac{1-d\psi_1}{1-(b+d)\psi_1}\right\}^\nu \\ &= \phi_\nu(b=c=0)\{1+\sum_{k=1}^{\infty}[\sum_{n=0}^{k}\binom{-\nu}{k-n}\binom{\nu}{n}(b+d)^{k-n}d^n](-\psi_1)^k\} \\ &= \phi_\nu(b=c=0)\{1+\sum_{k=1}^{\infty} h(b,k)\psi_1{}^k\} \end{aligned} \tag{2.40}$$

where $h(b,k)$ is defined in (2.38). Since the infinite series in (2.40) is uniformly convergent we can invert termwise to get (2.37). It is easily verified that b=0 leads to $g_\nu(x,y)$ in (2.26) and $\nu=1$ gives $f_1(x,y)$ in Corollary 1 with c=0.

As stated before, equation (2.13) characterizes a bivariate exponential distribution and for that case $\psi_1(t_1,0)\psi_2(0,t_2) = [1-\theta_1 it_1]^{-1}[1-\theta_2 it_2]^{-1}$ where θ_1 and θ_2 are different from those in (2.16). From the characterization, Paulson (1973) formed a bivariate gamma distribution and for the special case of c=0 a

corresponding result to Theorem 2 is

$$f_\nu(x,y) = g_\nu(x,y) + g_\nu(x,y) \overset{x}{*} \sum_{k=1}^{\infty} h(b,k)\left(\frac{1}{\theta_1}\right)^k \frac{x^{k-1}e^{-x/\theta_1}}{\Gamma(k)} \quad (2.41)$$

where

$$g_\nu(x,y) = \frac{a^\nu}{\theta_1\theta_2\Gamma(\nu)}\left(\frac{xy}{d\theta_1\theta_2}\right)^{\frac{1}{2}(\nu-1)} e^{-x/\theta_1 - y/\theta_2}\, I_{\nu-1}\left(2\left(\frac{dxy}{\theta_1\theta_2}\right)^{\frac{1}{2}}\right), \quad (2.42)$$

$h(b,k)$ is the same as in (2.38), and $I_{\nu-1}\left(2\left(\frac{dxy}{\theta_1\theta_2}\right)^{\frac{1}{2}}\right)$ is the modified Bessel function of the first kind and order ν-1. That this result is true follows directly from (2.40) since ψ_1 in this situation is $[1-\theta_1 it_1]^{-1}$. For ν=1 and c=0, (2.41) checks with a result by Kohberger (1975). In a slightly different form (2.42) is the bivariate gamma distribution obtained implicitly by Wicksell (1933) and explicitly by Kibble (1941). The bivariate exponential distribution defined by (2.13) will be discussed in more detail in Part 4. We continue with some more properties of the BNB(a,b,c,p,q,ν) distribution.

By using the characteristic function $\phi_\nu(T)$ in (2.31) and the usual differentiation techniques there follows, after several tedious operations,

<u>Theorem 3:</u> The mean vector and covariance matrix for the BNB(a,b,c,p,q,ν) distribution are

$$\underline{\mu} = \begin{bmatrix} \nu\theta_1 \\ \nu\theta_2 \end{bmatrix} \quad (2.43a)$$

and

$$\Sigma = \begin{bmatrix} \nu\theta_1(1+\theta_1) & \frac{\nu(ad-bc)}{1-d}\theta_1\theta_2 \\ \frac{\nu(ad-bc)}{1-d}\theta_1\theta_2 & \nu\theta_2(1+\theta_2) \end{bmatrix}. \tag{2.43b}$$

We digress briefly at this point and give a method of moments parameter estimation technique for the special case b=c=0.

1. In σ_X^2 and σ_Y^2 set $\theta_1=\bar{x}/\nu$ and $\theta_2=\bar{y}/\nu$ and take the product of σ_X^2 and σ_Y^2 to be equal to the product of the sample variances. There results a quadratic function in ν:

$$(1 - \frac{s_X^2 s_Y^2}{\bar{x}\bar{y}})\, \nu^2 + (\bar{x}+\bar{y})\nu + \bar{x}\bar{y} = 0 \tag{2.43c}$$

and for $s_X^2 s_Y^2 > \bar{x}\bar{y}$, which is expected if the marginals are approximately negative binomial, there is exactly one positive root that we can take as our estimate of ν.

2. In σ_{XY} substitute $\theta_1=\bar{x}/\nu$ and $\theta_2=\bar{y}/\nu$ and set σ_{XY} equal to the sample covariance, s_{XY}. Solving for the parameter a, we have

$$a = 1 - \frac{\nu s_{XY}}{\bar{x}\bar{y}}. \tag{2.43d}$$

3. From (2.16), $\theta_1=p/[(1-p)a]$ and so

$$p = \frac{a\theta_1}{1+a\theta_1} = \frac{a\bar{x}/\nu}{1+a\bar{x}/\nu}. \tag{2.43e}$$

4. In a like manner,

$$q = \frac{a\bar{y}/\nu}{1+a\bar{y}/\nu}. \tag{2.43f}$$

The next theorem establishes that the regression function for the bivariate geometric distribution in Corollary 1 is non-linear.

Theorem 4: For the BNB(a,b,c,p,q,1) distribution (bivariate geometric of Paulson and Uppuluri) where $b \neq 0$ the regression function is

$$E[Y|x] = \frac{q}{1-q}\left[\frac{b+d}{b} + \left(\frac{A}{m} - \frac{d}{b}\right)k^{x+1}\right] \qquad (2.44)$$

where

$$m = p + (1-p)(a+c),\ A = cm/[(a+c)(a+b)],$$

$$k = m/[m + b(1-p)] \text{ and } d = 1-a-b-c. \qquad (2.45)$$

Proof: For the conditional mean $E[Y|x]$, which is a function of x, take as a generating function the z-transform

$$\mathcal{Z}(E[Y|x]) = g(z) = \sum_{x=0}^{\infty} z^{-x}\, E[Y|x]. \qquad (2.46)$$

By using the definition of $E[Y|x]$ we can write

$$g(z) = (1+\theta_1) \sum_{x=0}^{\infty} \sum_{y=0}^{\infty} \left[\frac{(1+\theta_1)}{z\theta_1}\right]^x y f_1(y,x) \qquad (2.47)$$

since the marginal density of X is $\Pr[X=x] = \left(\frac{1}{1+\theta_1}\right)\left(\frac{\theta_1}{1+\theta_1}\right)^x$. This last equation can be written as

$$g(z) = (1+\theta_1) \frac{\partial}{\partial\lambda} P\left[\frac{1+\theta_1}{z\theta_1},\lambda\right]\Big|_{\lambda = 1} \qquad (2.48)$$

where $P[\cdot,\cdot]$ is the probability generating function of the bivariate pair (X,Y). The probability generating function can be obtained from $\phi_\nu(T)$ by substituting $r=e^{it_1}$ and $s=e^{it_2}$ and after some computation and simplification we get

$$g(z) = \left(\frac{q}{1-q}\right)\left(\frac{1}{m+b(1-p)}\right)\left(\frac{Az}{z-k} + \frac{z(z-m)}{(z-k)(z-1)}\right) \qquad (2.49)$$

where m, k and A are defined in (2.45). Inverting g(z) (Jury (1964)) gives (2.44). This type of a nonlinear regression function is sometimes called an exponential regression function.

In a direct way using the definition of $E[Y|x]$ we obtain

Theorem 5: For the BNB(a,0,0,p,q,ν) distribution

$$E[Y|x] = \frac{q}{1-q}\left[\frac{\nu+d(1-p)x}{1-d(1-p)}\right] \qquad (2.50)$$

where d=1-a.

The following theorem is very useful for computations.

Theorem 6: For the BNB(a,0,0,p,q,ν) distribution the probability function can be computed with the recursive formula

$$g_\nu(x,y+1) = \frac{q}{(y+1)(1-z)}[(x+y+\nu)g_\nu(x,y) - p(\nu+x-1)g_\nu(x-1,y)], \qquad (2.51)$$

$x \geq 1$, $y \geq 0$ and $z = d(1-p)(1-q)$. Coupled with

$$g_\nu(0,0) = \left[\frac{a(1-p)(1-q)}{1-z}\right]^\nu$$

the probability function is determined.

Proof: By using the following relation for $F(a,b;c;z)$ and two of its contiguous functions (15.2.18, Abramowitz and Stegun (1964)):

$$(c-a-b)F(a,b;c;z) - (c-a)F(a-1,b;c;z)$$
$$+ b(1-z)F(a,b+1;c;z) = 0,$$

(2.26) can be written as

$$g_\nu(x,y) = a^\nu h_1(x)h_2(y)\{\frac{1}{x+y+\nu}[xF(\nu+x-1,\nu+y;\nu;z)$$
$$+ (\nu+y)(1-z)F(\nu+x,\nu+y+1;\nu;z)]\}$$

and (2.51) follows easily with some elementary operations. The cell probability at (0,0) is obvious from (2.26) and (2.23).

An important result for parameter estimation is next.

<u>Theorem 7</u>: For the $\text{BNB}(a,0,0,p,q,\nu)$ distribution and if ν is known the likelihood equations for a random sample of size n are

$$\frac{\partial \log L}{\partial a}: \quad \frac{\nu}{a} + \bar{y} - \bar{R} = 0 \qquad (2.52a)$$
$$\frac{\partial \log L}{\partial p}: \quad \frac{(1-p)}{p}\bar{x} + \bar{y} - \bar{R} = 0 \qquad (2.52b)$$
$$\frac{\partial \log L}{\partial q}: \quad \frac{(1-q)\bar{y}}{q} + \bar{y} - \bar{R} = 0 \qquad (2.52c)$$

where L is the likelihood function, $\bar{x}$ and $\bar{y}$ are the sample means for the marginal distributions, $\bar{R}=\frac{1}{n}\sum_{x,y} n_{xy}(\frac{y+1}{q})\frac{g_\nu(x,y+1)}{g_\nu(x,y)}$,

n_{xy} is the number of observations for which X=x, Y=y, and $g_\nu(x,y)$ is the probability function in (2.26).

Proof: If the probability function in (2.26) is differentiated with respect to the parameters a, p and q the following differential-difference equations result:

$$\frac{\partial g_\nu(x,y)}{\partial a} = \left(\frac{\nu}{ad} + \frac{y}{d}\right) g_\nu(x,y) - \frac{1}{d}\left(\frac{y+1}{q}\right) g_\nu(x,y+1) \qquad (2.53a)$$

$$\frac{\partial g_\nu(x,y)}{\partial p} = \left(\frac{x}{p} + \frac{y}{1-p}\right) g_\nu(x,y) - \frac{1}{1-p}\left(\frac{y+1}{q}\right) g_\nu(x,y+1) \qquad (2.53b)$$

$$\frac{\partial g_\nu(x,y)}{\partial q} = \left(\frac{y}{q} + \frac{y}{1-q}\right) g_\nu(x,y) - \frac{1}{1-q}\left(\frac{y+1}{q}\right) g_\nu(x,y+1). \qquad (2.53c)$$

These equations follow by using (15.2.1, Abramowitz and Stegun)

$$\frac{\partial F(a,b;c;z)}{\partial z} = \frac{ab}{c} F(a+1,b+1;c+1;z)$$

and (exercise 1, page 296, Whittaker and Watson (1965))

$$F(a,b+1;c;z) - F(a,b;c;z) = \frac{az}{c} F(a+1,b+1;c+1;z).$$

The log likelihood function, log L, for a random sample of size n is $\sum_{x,y} n_{xy} \log g_\nu(x,y)$ and so

$$\frac{\partial \log L}{\partial a} = \sum_{x,y} n_{xy} \frac{1}{g_\nu(x,y)} \frac{\partial g_\nu(x,y)}{\partial a} .$$

Using (2.53a) and a few simple operations leads to (2.52a). Similarly, (2.52b) and (2.52c) obtain. We were unable to get a differential-difference equation involving $\frac{\partial g_\nu(x,y)}{\partial \nu}$, ν assumed unknown.

From (2.52) it is clear that

$$\frac{(1-p)}{p}\bar{x} = \frac{(1-q)}{q}\bar{y} = \frac{\nu}{a} \tag{2.54}$$

and these relationships are very useful in estimating the parameters via the method of maximum likelihood. For ν known or not, the conditions in (2.52) are necessary for a maximum likelihood solution to the likelihood function for $b=c=0$. Therefore, (2.54) can be used to reduce the dimensionality of the unknown parameter space from four if ν is unknown, to two by taking, say, $p = q\bar{x}/[q\bar{x}+(1-q)\bar{y}]$ and $a = \nu q/[(1-q)\bar{y}]$. We have used a nonlinear optimization computer program (Cross (1970)) to solve for the parameter estimates and the dimensionality reduction permits extremely shorter running times.

Next we show a corresponding result to Theorem 7 for Paulson's bivariate gamma distribution. For the distribution defined in (2.42) and if ν is known the likelihood equations for a random sample of size n are:

$$\begin{array}{ll} \dfrac{\partial \log L}{\partial a}: & \left\{ \begin{array}{l} \dfrac{\nu}{a} - \dfrac{\nu+1}{2} - \bar{S} = 0 \\[2ex] \dfrac{\bar{x}}{\theta_1} - \dfrac{\nu+1}{2} - \bar{S} = 0 \\[2ex] \dfrac{\bar{y}}{\theta_2} - \dfrac{\nu+1}{2} - \bar{S} = 0 \end{array} \right. \\ \dfrac{\partial \log L}{\partial \theta_1}: & \\ \dfrac{\partial \log L}{\partial \theta_2}: & \end{array} \tag{2.55}$$

where L, $\bar{x}$ and $\bar{y}$ are as in Theorem 7, $a=1-d$, $\bar{S}=\frac{1}{n}\sum_{x,y} I^*/I$, $I^*=(z/2)\frac{\partial I_{\nu-1}(z)}{\partial z}$, $I=I_{\nu-1}(z)$ and $z=2[(1-a)xy/(\theta_1\theta_2)]^{\frac{1}{2}}$. The result follows, after some lengthy but straightforward computa-

tions, by taking $L = \prod_{x,y} g_\nu(x,y)$ where $g_\nu(x,y)$ is in (2.42) and forming

$$\frac{\partial \log L}{\partial \lambda} = \sum_{x,y} \frac{1}{g_\nu(x,y)} \frac{\partial g_\nu(x,y)}{\partial \lambda}$$

for $\lambda \varepsilon \{a, \theta_1, \theta_2\}$. It is obvious that necessary conditions for the parameters are

$$\frac{\bar{x}}{\theta_1} = \frac{\bar{y}}{\theta_2} = \frac{\nu}{a} \; . \tag{2.56}$$

We show in a later section the bnb of this section fitted to some data. First we introduce another bnb distribution which has certain desirable properties.

2.3 A New Bivariate Negative Binomial Distribution, Via a Trivariate Reduction, and Some Properties

The two previously discussed bivariate negative binomial distributions have marginals which are negative binomial with parameters θ_i, i=1,2, and common parameter ν. In this section we introduce another bivariate negative binomial distribution whose marginals have parameters ν_i, i=1,2, and common parameter θ. Data are shown later for which this latter model seems more appropriate.

We construct via reduction of a certain trivariate negative binomial distribution with independent marginals a bivariate negative binomial distribution. Mardia (1970) refers to this as a trivariate reduction; Holgate (1964) used this technique to construct a bivariate Poisson distribution and Arnold

(1967) generalized the procedure.

Theorem 8: Let X_1, X_2, X_3 be independent negative binomial r.v.'s with common parameter θ and individual parameters ν_1, ν_2 and ν_3, respectively. Then the probability function of $X = X_1 + X_2$, $Y = X_2 + X_3$ is given by

$$h(x,y) = \left(\frac{1}{1+\theta}\right)^{\nu_1+\nu_2+\nu_3}\left(\frac{\theta}{1+\theta}\right)^{y} \sum_{w=k}^{x} \binom{\nu_1+w-1}{w} \times$$

$$\binom{\nu_2+x-w-1}{x-w}\binom{\nu_3+y-x+w-1}{y-x+w}\left(\frac{\theta}{1+\theta}\right)^{w} \qquad (2.57)$$

where

$$k = \begin{cases} 0, & \text{if } x \leq y \\ x-y, & \text{if } x > y. \end{cases}$$

Proof: The joint distribution of X_1, X_2, X_3 is

$$f(x_1,x_2,x_3) = \left(\frac{1}{1+\theta}\right)^{\nu_1+\nu_2+\nu_3} \prod_{i=1}^{3} \binom{\nu_i+x_i-1}{x_i}\left(\frac{\theta}{1+\theta}\right)^{x_i}$$

and by taking the transformation of variables $X = X_1 + X_2$, $Y = X_2 + X_3$ and $W = X_1$ it follows in the usual way that the joint distribution of (X,Y) is (2.57).

This bivariate negative binomial distribution is designated BNB-TR$(\theta,\nu_1,\nu_2,\nu_3)$ and from the defining relations its marginals are negative binomial; X has parameters θ and $\nu_1 + \nu_2$ and Y has parameters θ and $\nu_2 + \nu_3$ (Johnson and Kotz). The marginal means

and variances are thus known and the covariance of (X,Y) is

$$\operatorname{cov}(X,Y) = \operatorname{cov}(X_1+X_2,\ X_2+X_3) = \sigma^2_{X_2}. \tag{2.58}$$

Therefore, $\rho = \nu_2/[(\nu_1+\nu_2)(\nu_2+\nu_3)]^{\frac{1}{2}}$ and we see that ρ is restricted to nonnegative values.

<u>Theorem 9</u>: The regression function for the bivariate r.v. in (2.57) is

$$E[Y|x] = \nu_3\theta + \frac{\nu_2 x}{\nu_1+\nu_2}. \tag{2.59}$$

Proof: Given X=x, the r.v. $Y|x$ has expectation $E[X_2|x] + \nu_3\theta$ since X_3 and X are independent and so we desire the distribution of $X_2|x$. By writing the joint distribution of X_1 and X_2 and transforming to new variables by letting $X = X_1 + X_2$, $X_2 = X_2$, we obtain for the joint distribution of (X,X_2),

$$f(x,x_2) = \binom{\nu_1+x-x_2-1}{x-x_2}\binom{\nu_2+x_2-1}{x_2}\left(\frac{1}{1+\theta}\right)^{\nu_1+\nu_2}\left(\frac{\theta}{1+\theta}\right)^x,$$

$$0 \le x_2 \le x. \tag{2.60}$$

From definitions it follows that

$$f(x_2|x) = \frac{\binom{\nu_1+x-x_2-1}{x-x_2}\binom{\nu_2+x_2-1}{x_2}}{\binom{\nu_1+\nu_2+x-1}{x}},\quad 0\le x_2\le x \tag{2.61}$$

and

$$E[X_2|x] = \frac{\nu_2 x}{\nu_1+\nu_2}. \tag{2.62}$$

Equation (2.59) results.

The next section shows how these bnb distributions compare with the Guldberg-Bates-Neyman model in fitting bivariate data related to shift worker absenteeism and disease data for industrial workers.

2.4 Bivariate Negative Binomial Distributions Fitted to Data

In this section we fit two data sets from the literature with the three previously discussed bnb distributions. Our objective in using these data is to illustrate certain aspects of the distributions. The first data set is given by Arbous and Sichel (1954) and concerns absenteeism for 248 shift workers in two adjacent yearly time periods and the second one is due to Bates and Neyman (1952) and shows the number of cases of incapacity suffered, per individual, during a common time period and due to two diseases; the sample size is 1286. For the absenteeism data we show that either the G-B-N(α,θ,ν) model in (2.6) or the BNB$(a,0,0,p,q,\nu)$ model in (2.26) fit the bivariate data reasonably well but that the regression function from the latter model describes better the observed conditional means. None of the three models adequately describe the joint distribution of the disease data even though the marginals are acceptably fitted by a univariate negative binomial distribution. It

will be shown, however, that the two new bnb distributions do describe certain properties of the disease data better than the G-B-N(α,θ,ν) distribution, though.

Parameter estimates for the various models are by the method of maximum likelihood estimation (MLE). Bates and Neyman show how to solve the likelihood equations for the G-B-N(α,θ,ν) distribution; for the BNB(a,b,c,p,q,ν) and BNB-TR (θ,ν_1,ν_2,ν_3) distributions we use a computer optimization program (Cross) directly on the log likelihood function to obtain our estimates. For the parameters b and c defined to be zero in the BNB(a,b,c,p,q,ν) model we use the necessary conditions in (2.54) to reduce the dimensionality of the parameter space.

Corresponding to the observed pairs (x_i,y_i), $i=1,2,\ldots,n$, representing a random sample from the unknown probability function $f(x,y)$, we wish to test the hypothesis H_0: $f(x,y)=f_0(x,y)$, where $f_0(x,y)$ is specified to be one of the referenced bnb distributions. The χ^2 test is used as a goodness-of-fit test of H_0; we point out that the problems with grouping cells which are generally associated with this test in univariate settings are even more dramatic for bivariate cases. To illustrate, Table 10 shows two independent groupings for a data set to be described in the next part. Before commenting on the table, certain remarks are required.

Following the practice of Bates and Neyman in their paper, cells which have expected frequencies less than three are

grouped. In Table 10 and similar ones to follow, the dashed lines indicate the boundary of the particular cells; heavy lines indicate the grouping adopted for the application of the χ^2 test. Three numbers are shown in each cell: the observed frequencies are shown in the upper left corner of particular cells and the decimal numbers are the expected frequencies on the left and the contributions to χ^2 on the right. If several adjoining cells are grouped then the expected frequency and χ^2 values shown are for the entire group. The P value given is the probability of obtaining a value of χ^2 exceeding the computed amount assuming H_0 is true.

Table 10 clearly shows how the probability P is affected by different groupings. Unlike the univariate situation we have two directions to contend with here for grouping and it is not obvious how to proceed. This is pointed out to emphasize the need for an alternate goodness-of-fit test for bivariate data (and in general multivariate data), one perhaps being independent of any grouping. For lack of a better test we resort to the χ^2 test. In every fit to be shown herein the groupings are completely independent of the observations and the findings are from a single attempt at grouping the expected frequencies for the minimum value of three.

Table 1 shows certain summary results for all of the data to be analyzed. The first column identifies the data, column 2 gives the sample size and correlation, column 3 specifies the marginal random variables and shows the associated sample means

and variances, columns 4 and 5 show the parameters of the univariate negative binomial fitted to the marginals and the associated χ^2 values, degrees of freedom and probability levels, respectively. Results from this table will be presented along with a discussion of the individual data sets. The absenteeism data is examined first.

One of the objectives of the Arbous-Sichel paper was to extend the notion of accident-proneness introduced by earlier workers (see Kemp (1970) for a history) to absence-proneness in shift workers. Table 2 shows the observed and expected cell frequencies for the number of absences in two adjacent yearly time periods (1947 and 1948) for 248 workers; the expectations are from the G-B-N$(1,\theta,\nu)$ symmetric model $(\alpha=1)$.

From Table 1 we see that the marginal distributions are fitted rather nicely by the univariate negative binomial and coupled with the fairly large sample correlation coefficient, it seems reasonable to expect a bnb distribution to adequately describe the data.

Table 3 shows the expected cell frequencies from the BNB $(a,0,0,p,q,\nu)$ distribution; no goodness-of-fit test is attempted for this data set since Arbous and Sichel do not show their grouping. They report a χ^2 value of 17.0 with 13 degrees of freedom (df) and $P \doteq 0.20$ associated with the G-B-N$(1,\theta,\nu)$ model so certainly the fit is reasonable. A visual comparison of the expected cell frequencies for the two models indicates a

close agreement and so we would expect a similar probability P to obtain for the fit of the BNB$(a,0,0,p,q,\nu)$ distribution.

Although the fit of the G-B-N$(1,\theta,\nu)$ model to the observed data is reasonably good, the authors point out that 12 of 18 observed means lie below the theoretical regression function (MLE estimates are used in (2.12)). The BNB$(a,0,0,p,q,\nu)$ model, and using MLE estimates, gives rise to a regression function for which only 10 of the 18 observed means are less than the predicted values.

We tried to fit the BNB$(a,b,c,p,q,1)$ and the BNB$(a,b,0,p,q,\nu)$ models in equations (2.36) and (2.37), respectively, to these data but got zero MLE estimates of b and c in the first model and an estimate of b equals zero in the second. The apparent lack of influence of the parameters b and c will be noted again for the Bates-Neyman data. No attempt was made to fit the BNB-TR $(\theta,\nu_1,\nu_2,\nu_3)$ distribution of (2.57) to this data set since the marginal estimates of the parameter ν (column 4 of Table 1) are approximately the same. Next we discuss the disease data.

In their paper Bates and Neyman present several data sets related to injuries and diseases suffered during a common period of time by office and industrial workers. For the set associated with two kinds of diseases for 1286 industrial workers the fit of the G-B-N(α,θ,ν) distribution is deficient both in describing the bivariate data and the observed regression function. That a bnb distribution should be a candidate model for the

data, though, is suggested by Table 1 since the marginals are reasonably well fitted by a univariate negative binomial and the sample correlation coefficient is moderate.

In the following diagram we show the empirical and theoretical regression functions for this data; theoretical values result by substituting MLE parameter estimates, via fitting the joint raw data, into (2.12). Although a "by-eye" fit of a regression function to the data fails somewhat due to the fact that the observed means are based on varying sample sizes, shown at the bottom of the diagram, it is apparent that there is some nonlinearity in the data.

THEORETICAL AND OBSERVED REGRESSION FUNCTIONS FOR BATES-NEYMAN DATA

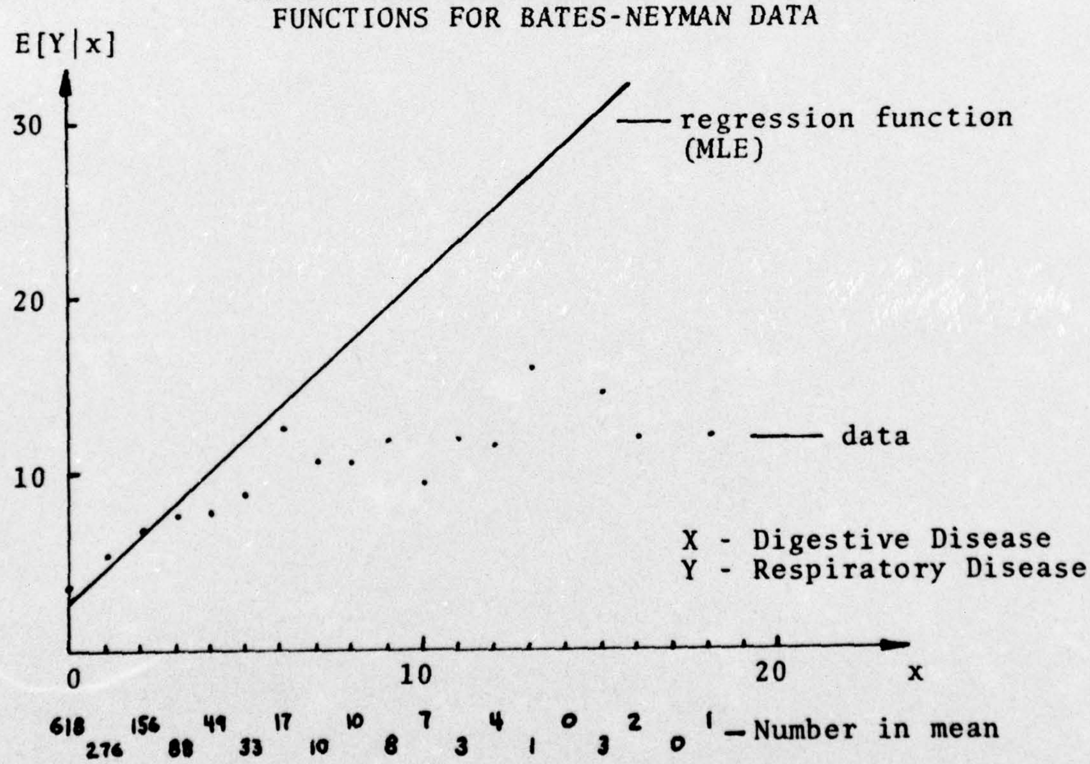


Table 4 shows the observed and expected cell frequencies given by Bates and Neyman for the disease data; the expected frequencies are based upon the G-B-N(α,θ,ν) distribution. From the previous diagram and the table it is clear that this particular model fails to adequately describe the joint data and the regression function so it is natural to seek a better representation for the data. To this end we fit to the data the new bnb distributions; our results are not completely successful in that the joint fits are also inadequate even though the χ^2 values are much reduced from that of the G-B-N(α,θ,ν) fit. That the empirical regression function is described better will be illustrated.

We now give in displays like Table 4 the results of applying the new bnb distribution to these data. In order to provide a close comparison with the Bates-Neyman results we change their groupings only when necessary to maintain the minimum expectation of three. Tables 5 through 8 show the results and Figure 1 gives some of the associated regression functions.

Table 5 illustrates our first attempt at fitting these data; the bivariate geometric was chosen since it leads to a nonlinear regression function (see 2.44). Choosing a criterion of χ^2, the fit is much better than the one in Table 4 but still inadequate. Figure 1a gives the regression function from (2.44) with MLE estimates; 1b shows the least squares fit of the curve $\alpha_0+\alpha_1 k^x$, which is the form of (2.44), to the data. We emphasize the

fact that all parameter estimates are based on the raw data; that is, these curves were not fitted to the illustrated means.

Tables 6 and 7 show how the BNB$(a,b,0,p,q,\nu)$ in (2.37) and the BNB$(a,0,0,p,q,\nu)$ in (2.26), respectively, describe the bivariate data. Figures 1c and d give the corresponding regression functions. We note that the regression function for the BNB$(a,b,0,p,q,\nu)$ distribution is nonlinear. It is observed that the χ^2 values are approximately the same in Tables 5, 6 and 7 so none of the special cases considered seem superior in describing the joint observations. The regression functions are different, though, and it appears as if the BNB$(a,b,0,p,q,\nu)$ regression model is best, at least among the ones based on MLE estimates.

A possible reason for the difficulty in adequately fitting these data is that each of the distributions discussed in Tables 4 through 7 has a common ν parameter associated with the marginals, but the individual sample values of ν from Table 1 are quite different. For one marginal the estimate of ν is 0.53 and for the other, 1.69. Thus we are lead to apply the BNB-TR $(\theta,\nu_1,\nu_2,\nu_3)$ distribution in (2.57); Table 8 displays the fit. We see an improvement in the χ^2 value but still not enough to produce a reasonable fit. Figure 1e shows the regression function for this model. For reference we give the least squares linear regression in Figure 1f.

Associated with each bivariate display in Tables 4 to 8 is an implied marginal fit. None of the distributions adequately describe both observed marginal distributions although the G-B-N(α,θ,ν) model does describe adequately ($P\doteq0.30$) the random variable labeled respiratory disease and the BNB-TR $(\theta,\nu_1,\nu_2,\nu_3)$ model gives $P\doteq0.05$ for the digestive disease marginal and $P\doteq0.02$ for the other one.

For the Arbous-Sichel data and for special cases of the general BNB(a,b,c,p,q,ν) model we saw that the parameters b and c were not needed to describe those data. Here for the Bates-Neyman data we see from Tables 6 and 7 that the impact of the parameter b appears minimal in that the expected cell frequencies are about the same. Ignoring the observation that the BNB$(a,b,0,p,q,\ )$ model leads to a nonlinear regression function, whereas the BNB$(a,0,0,p,q,\nu)$ distribution gives a linear one as shown in (2.50), we suspect for parent populations with positive correlation that the latter model is fairly robust against alternatives involving nonzero parameters b and c. Although not attempted here, this conjecture could be examined via simulation; plots of the probability surface as a function of some of the parameters could be helpful too.

2.5 Summary

In this part we discussed the bnb distribution introduced by Guldberg (1934) and generalized by Bates and Neyman (1952). This bnb distribution admits of positive correlations and linear

regression functions. Also we introduced and derived several properties for two new bnb distributions, one obtained by convoluting a bivariate geometric distribution given by Paulson and Uppuluri (1972), and another obtained by reducing a certain trivariate negative binomial distribution. For the convolution process a dual bivariate gamma distribution exists (Paulson (1973)) and for it the duality implies the exactly analogous properties.

For the bnb distribution resulting from a convolution, labeled BNB(a,b,c,p,q,ν), we established the following results:

(1) for ν integer, the probability function in (2.32),

(2) for $\nu>0$, the moments to order two in (2.43 a and b),

(3) for $c=0$ and $\nu>0$, the probability function in (2.37), the probability density function for the dual bivariate gamma distribution in (2.41) and the nonlinearity of the regression function in Figure 1c,

(4) for $\nu=1$, the probability function in (2.36) and the equation for the regression function (nonlinear) in (2.44),

(5) for $b=c=0$ and $\nu>0$, a method of moments parameter estimation technique in (2.43 c-f), the equation for the regression function in (2.50), a recursive formula for the probabilities in (2.51), the likelihood equations for a random sample (assume ν known) in (2.52), the differential-difference equations (ν known) in (2.53), in (2.54) a necessary relationship for the parameters in optimizing the likelihood function and

the likelihood equations for the dual bivariate gamma in (2.55) and the necessity condition for its parameters in (2.56).

Contrary to the Guldberg-Bates-Neyman distribution this one gives rise to positive or negative correlations and to linear or nonlinear regression functions.

For the bnb distribution which resulted from a reduction we derived the probability function in (2.57), the moments to order two in (2.58) and the regression function in (2.59). This distribution is basically different from the two previously discussed ones in that its marginals have characteristic functions of the form $[1+\theta(1-e^{it})]^{-\nu_i}$, i=1,2, whereas the characteristic functions associated with the latter distributions are of the form $[1+\theta_i(1-e^{it})]^{-\nu}$, i=1,2. Equation (2.1) shows how the resulting marginal distributions would differ. From the applications of these distributions to data we suspect that bnb models which give rise to marginals with characteristic functions of the form $[1+\theta_i(1-e^{it})]^{-\nu_i}$, i=1,2, would be useful.

We applied these distributions to the bivariate data given by Arbous and Sichel (1954) on absenteeism among 248 shift workers in two yearly periods and to disease data of two types among 1286 industrial workers.

Arbous and Sichel fitted the symmetric (α=1) Guldberg-Bates-Neyman model in (2.6) to absenteeism data and got a reasonable fit ($P \doteq 0.20$) but the regression function, with parameters via MLE, overestimated the observations in that 12 of

18 means were below the computed regressions. We applied the BNB$(a,0,0,p,q,\nu)$ distribution to these joint data and the regression function was such that only 10 of 18 means were below it. Arbous and Sichel did not show the cell groupings they adopted for use in the χ^2 test for the bivariate fit so we could not compare the BNB$(a,0,0,p,q,\nu)$ model to the Guldberg-Bates-Neyman model. By-eye, the fits seemed comparable.

Bates and Neyman applied the Guldberg-Bates-Neyman distribution to disease data but the fit did not adequately describe the observed bivariate data or the observed means which clearly suggested a nonlinear form for the regression function. We applied several special cases of the BNB(a,b,c,p,q,ν) model and the so called BNB-TR$(\theta,\nu_1,\nu_2,\nu_3)$ model in (2.57) to these data and never got a reasonable fit although the values of χ^2 were much reduced; from the latter model the χ^2 value was about one-half of the value reported by Bates and Neyman. The nonlinear regression functions resulting from the special cases of the BNB(a,b,c,p,q,ν) distribution described much better the observed means.

From our experience in fitting special cases of the BNB (a,b,c,p,q,ν) distribution to these data and to the data to be discussed in the next part we suspect that the parameters b and c are relatively unimportant in fitting bivariate data from populations with positive correlation. Except in those

cases where the empirical regression function is obviously nonlinear the BNB$(a,0,0,p,q,\nu)$ case is probably a fairly robust model. The recursive formula and parameter reduction technique developed here make it relatively easy to work with, too.

PART 3

SOME BIVARIATE APPROACHES FOR ANALYZING AIRCRAFT OPERATIONS AND MAINTENANCE DATA

3.1 Introduction and Historical Review

Our objective in this part is to apply bivariate distributions to some problems related to inventory control and maintenance in military aircraft logistics. In a way to be shown we form bivariate r.v.'s related to these problems and illustrate their utility with actual data. Although the discussion is restricted to applying two of the aforementioned bnb distributions to certain data sets, the techniques are applicable to other settings.

The analyses presented here, for the most part, are in the context of fitting distributions to bivariate data and then using these sample distributions to address certain problems. Several authors have postulated areas in reliability where certain continuous bivariate distributions can be expected to result. See, for example, Downton (1970), Harris (1968), Hawkes (1972), and Marshall and Olkin (1967) who study bivariate exponential distributions. Closer to the techniques envisioned here are the works of Fawcett and Gilbert (1966) for characterizing demand patterns (univariate) for aircraft spare parts and Youngs, Geisler and Brown (1955) for predicting demand for aircraft spare parts using the method of conditional probabilities. The account by Dade (1973) for some alternative approaches to maintenance analysis is of interest also.

In the next section we present some bivariate approaches to problems dealing with discrete data. Section 3 describes how similar approaches could be used with continuous data.

3.2 Some Bivariate Analyses of Discrete Data

The purpose of this section is to illustrate with several examples how discrete bivariate r.v.'s can be associated with certain aircraft operations and maintenance problems and subsequently give rise to meaningful analysis techniques. In particular, we present demand data for aircraft spare parts and use bivariate distributions to suggest how a particular kind of inventory model can be constructed; additionally, aircraft abort data are given and it is shown how the regression function for a certain bivariate r.v. related to these data can be used to suggest the effect of overhaul on an aircraft's performance. Besides the above applications an important observation in its own right is that these data, properly defined, can be described adequately by univariate and bivariate negative binomial distributions.

First we show that bnb distributions adequately describe demand data for aircraft spare parts in two adjacent time periods. Table 9a gives for a random sample of 72 aircraft parts actual demand data for a four month period where we have formed bivariate data by splitting the period into two smaller intervals; we take the first interval to be the first three months and the second interval to be the fourth month. As an illustra-

tion of the data two parts were demanded five times in the first three months and then a single time in the fourth month. From Table 1 we see that the sample correlation coefficient for the bivariate data is 0.54 and that the marginals are fitted, reasonably well, with the univariate negative binomial distribution (illustrated in 9b). We expect then that one of the bnb distributions should describe these data and in Table 9a we show the fit of the Guldberg-Bates-Neyman model of (2.6). (In this entire part procedures for estimating parameters, applying the χ^2 test and illustrating results are all the same as were described in Section 2.4.) For this fit we have $P \doteq 0.18$ indicating that approximately 18 of 100 fits would be worse, assuming, of course, that the G-B-N(α,θ,ν) model is the underlying parent population. That another bnb distribution describes these data is illustrated next.

In Table 10 we show how the BNB($a,0,0,p,q,\nu$) distribution in equation (2.26) describes the data; this is the same data set that was used before to illustrate how the P value associated with the χ^2 test is affected by cell groupings. We reiterate that an alternate goodness-of-fit test, one, perhaps, being independent of cell groupings, would be desirable but here we use the χ^2 test and agree to report results based upon a single attempt at grouping the cells for the minimum expectation of three.

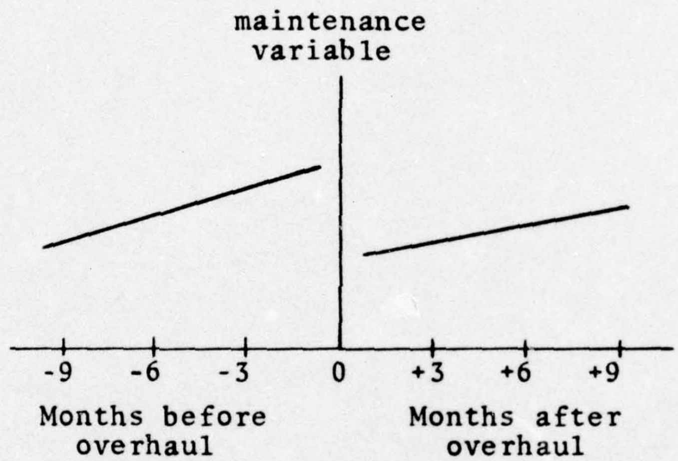
By inspection the fits illustrated in Tables 9 and 10 seem to be about the same. The P values associated with the marginal fits of the BNB($a,0,0,p,q,\nu$) distribution to the univariate ob-

servations are 0.24 and 0.27 and the G-B-N(α,θ,ν) model gives rise to similar values.

That bnb distributions can be applied to these kinds of data should be useful in inventory control and particularly to problems where short range predictions are required, such as in constructing fly-away kits. In military air operations certain parts are set aside from normal operating channels and in an emergency are supposed to provide enough spares to last for a fixed amount of time, usually one month. These parts make up a so called fly-away kit. Youngs, et. al. (1955) suggest this application in their report but show no distributions fitted to bivariate data as we do. We leave this area and next discuss some aircraft abort data.

That bivariate distributions which admit of negative correlations can be useful in applications is illustrated with the following data set. We show in Table 11 flight aborts (missions interrupted during flight) for a random sample of 109 aircraft for two consecutive six month time periods. The flight aborts are limited to those caused by materiel failures. From Table 1 the sample correlation coefficient is -0.16 and we see that one of the marginals is very well fitted with the univariate negative binomial but the other fit is inadequate. The expected cell frequencies in Table 11 are from the BNB(a,b,c,p,q,1) distribution and a P value of 0.12 results.

Next we give other aircraft abort data and show an analysis which suggests the effect of overhaul on an aircraft's performance. Aircraft undergo large scale overhaul programs periodically and it is important to know if the programs are beneficial. Traditionally, these programs have been justified by the common assumption that the aircraft are restored to a better condition; the following diagram depicts one way this benefit is perceived.



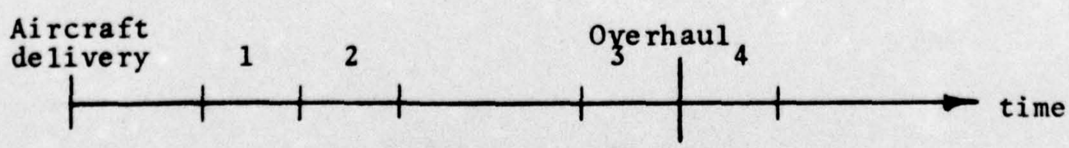
Before overhaul, for an individual aircraft, we show an increasing trend to denote degradation of a variable such as number of failures, number of aborts or perhaps, unscheduled maintenance manhours and an improvement immediately after the event followed by degradation again.

Previous attempts at investigating the effect of overhaul have centered on collecting sample data for some variable on either side of the event and then performing, say, an analysis of variance or a regression analysis to determine if such

degradation trends do, in fact, seem reasonable. Certain studies have indicated that aircraft are not in a degraded condition prior to overhaul and are not improved by the event (Dade (1973)). Some qualifying statements are required here--certainly an analysis of this nature is very complex in that many factors such as environment, previous missions, flying hour history and total age of the aircraft are involved so evidently no one analysis can be expected to be complete and exhaustive. These types of analyses can be suggestive of the effect of overhaul, though, and in this respect we seek to contribute another technique and illustrate, with some new data, its use. Although not attempted here common experimental design techniques could be employed to control some of these other factors.

We are primarily interested in developing techniques applicable to non-normal data and particularly to discrete data, such as aborts which are typically small. From here on we analyze total aborts, which are mission interruptions discovered during pre-flight or in-flight operations. As before we study only those aborts caused by materiel failures.

To examine the effect of overhaul we define two bivariate r.v.'s related to the periods shown in the following diagram. Periods 1 and 2 (common to all aircraft) are two adjacent six



month periods where no overhaul was performed and periods 3 and 4 are the six month periods immediately before and after an overhaul, respectively. The period lengths were chosen arbitrarily. If A_i is the number of total aborts per aircraft in period i, i=1,2,3,4, then we wish to compare the probability distribution of the pair (A_3,A_4) where an intervening overhaul was performed to the probability distribution of the pair (A_1, A_2) where no intervening overhaul occurred. Under a null hypothesis of the overhaul being ineffective in changing an aircraft's performance, as measured by aborts, these two distributions should be the same. The regression function is a descriptor of a bivariate distribution (Kendall and Stuart (1973)) and so one simple way to compare these two distributions would be to examine the two regression functions associated with sample data. All other factors being equal any two aircraft with the same number of aborts in periods 1 and 3 should have the same number of aborts, on the average, for periods 2 and 4 if overhaul is ineffective and so any difference in the corresponding regression functions should be suggestive of the effect of overhaul.

This type of analysis presupposes the existence of a suitable bivariate distribution which will adequately describe observed data. Next we present sample data related to the bivariate r.v.'s (A_1,A_2) and (A_3,A_4) and fit them with the BNB $(a,0,0,p,q,\nu)$ distribution in equation (2.26). This distribu-

tion was chosen arbitrarily and we suspect that other bnb distributions could be used equally well. Sample values of A_j will be denoted by a_j, $j=1,2,3,4$.

Table 12 gives the number of total aborts in periods 1 and 2 for 203 aircraft. To obtain this sample we considered an entire population of aircraft of a particular type (about 500) and excluded those aircraft with an overhaul during the time periods specified as 1 and 2 and also those aircraft without 12 full months of reported abort data during that time -- 203 aircraft resulted. As an illustration of the data three aircraft had ten total aborts in period 1 followed by six total aborts in period 2.

For these data we can use the ordinary sign test (Gibbons (1971)) for a bivariate r.v. to conclude that where no intervening overhaul is involved aircraft incur the same number of total aborts in two adjacent six month periods (here we are actually testing the hypothesis that the median of the r.v. (A_1-A_2) is zero versus a two-sided alternative; a normal approximation to the binomial used in this nonparametric test gives a sample value of $|z|=1.25$).

Table 1 shows certain descriptive statistics for these data and it is apparent that the univariate negative binomial distribution can be used to describe the marginal observations. We wish to describe further the bivariate data, though, and so we attempt to fit the joint observations with a bivariate negative binomial distribution.

Table 13 shows the expected cell frequencies which result by applying the BNB$(a,0,0,p,q,\nu)$ distribution of (2.26) to these data and Table 14 gives the observed and expected frequencies together and the χ^2 values. For convenience, we sum the observations in any one "super-cell" as is our custom for the expectations. We conclude that the bivariate fit, as measured by χ^2, is good ($P \doteq 0.65$). The P values for the associated marginal fits are 0.04 for A_1 and 0.56 for A_2. From equation (2.50) with MLE parameter estimates we see that the regression function, or the mean number of aborts for period 2 given the observed number of aborts for period 1, is $E[A_2|a_1]=4.53 + 0.26\ a_1$. It can be shown that the least squares regression function for these data is $E[A_2|a_1]=4.41 + 0.28\ a_1$. Next we show the same analysis for sample data from periods 3 and 4; that is, before and after an overhaul.

Tables 15, 16 and 17 give observations, expectations and the χ^2 test for periods 3 and 4 for 387 aircraft. These are the same type of aircraft as before and we have an aircraft being included in this sample if it has six full months of reported abort data on adjacent sides of a common overhaul event. We point out that periods 3 and 4 may be separated by two or three months which is usually the length of an overhaul for these aircraft.

If we apply the sign test to the bivariate data for (A_3, A_4) shown in Table 15, as was done for the observed data for

(A_1,A_2), there results a sample value of $|z|= 4.51$ and we can conclude that values of A_4 greater than A_3 are more likely in these data; that is, total aborts per aircraft are generally greater after overhaul.

Column 5 of Table 1 shows how the univariate negative binomial distribution fits the marginals. For the joint data, the P value (0.12) associated with the fit of the BNB(a,0,0, p,q,ν) distribution is not as high as for the previous data but we assume that it is acceptable. The implied marginal fits give rise to P values of 0.43 for A_3 and 0.05 for A_4. As a descriptor of these bivariate data, the regression function (using MLE estimates) is $E[A_4|a_3]=7.13 + 0.26\ a_3$. The least squares regression function is $E[A_4|a_3]=7.32 + 0.23\ a_3$. Figure 2 shows on one graph the estimated regression functions (via MLE) for these two data sets. We view the regression functions being useful in the following way. If two aircraft have the same number of aborts in periods 1 and 3, say, ten, then the aircraft which has an overhaul will have, on the average, about nine and one-half aborts in the next six months and the aircraft that does not get an overhaul will have about seven, again on the average. An important advantage in being able to fit the data as we have illustrated here is that a confidence interval could be placed on the predicted number of aborts by using the estimated bivariate probability function; thus, we avoid the normality assumption that is traditionally involved

with placing confidence limits on the predictions (Draper and Smith (1966)).

Apart from the development described here using bnb distributions and if we were willing to invoke the necessary normality assumptions, an alternative approach would be to use the least square regression functions in the above way. The results for these data would be about the same.

Of interest is another bivariate r.v. which is formed by taking A_4 and A_5 where A_5 is a r.v. representing the total aborts in the six months immediately after the fourth period. Thus, we have a new bivariate r.v. associated with the aircraft's performance right after overhaul and we are interested in how it compares with (A_3,A_4), the r.v. representing total aborts on either side of overhaul. Specifically, we are interested in whether or not aborts decrease again; if the overhaul is responsible for the observed increase in period 4 perhaps the degradation is similar to that usually experienced in a new item's performance and which gradually declines to a lower level.

Table 18 shows sample data for (A_4,A_5) from 428 aircraft of the same type previously considered; forming the differences $(a_4-a_5)_i$, $i=1,2,\ldots,428$, and using the sign test for an hypothesis of zero median difference for the r.v. A_4-A_5, versus a two-sided alternative, leads to a sample value of $z=0.74$ and so we take the hypothesis to be true for these data. Evidently

then, the number of total aborts is about the same for periods 4 and 5 for an individual aircraft.

We point out the rather large values of P, 0.75 for A_4 and 0.56 for A_5, which result from fitting the univariate negative binomial distribution to the marginal observations (see column 5 of Table 1). Although not illustrated, we applied to these sample data the same bnb distribution that was used in the two previous instances; the MLE parameter estimates which resulted are $\hat{a}=0.7485$, $\hat{p}=0.7728$, $\hat{q}=0.7642$ and $\hat{\nu}=2.0285$. Associated with the bivariate fit is $P \doteq 0.27$ and for the marginals, $P \doteq 0.66$ for A_4 and $P \doteq 0.76$ for A_5. We accept the fit. Based on MLE estimates, the regression function is $E[A_5|a_4]=6.97 + 0.20\ a_4$ and from least squares, $E[A_5|a_4]=6.84 + 0.21\ a_4$; the dashed line in Figure 2 shows the former. The upper two regression functions in the figure are probably within sampling error of one another; no attempt is made to test for true differences.

Although not done here it would be of interest to make a similar comparison before overhaul; that is, take a six month period preceding period 3 and form a bivariate r.v. for the number of total aborts in the new period and period 3 and then compare that r.v. to (A_3,A_4). Another analysis planned for the future is to compare (A_1,A_2) to (A_3,A_4) where just those aircraft with overhauls during periods 1 and 2 are selected for consideration in periods 3 and 4. In this way we have two samples during the same calendar time (approximately) and any

possible contributing causes from that factor would be somewhat controlled; Dade points out that changes in management policy or hardware and tactics connected to calendar time have often made maintenance data analyses difficult.

For a large group of aircraft of a particular type our results show how these aircraft performed after overhaul. The results are only suggestive; with only the surface analysis described here, we are not prepared to argue that the observations represent true degradation or that they were caused by the overhaul. We do think the approach is worthy of consideration though and coupled with a comprehensive experimental design could provide unbiased results.

Other definitions of the periods might lead to more revealing results; some alternatives are fixed flying hour intervals, a fixed number of sorties, or shorter monthly intervals. Here the total aborts were taken for the whole aircraft; perhaps total aborts for a particular subsystem would be better.

In the next section we discuss briefly similar applications for problems dealing with continuous data.

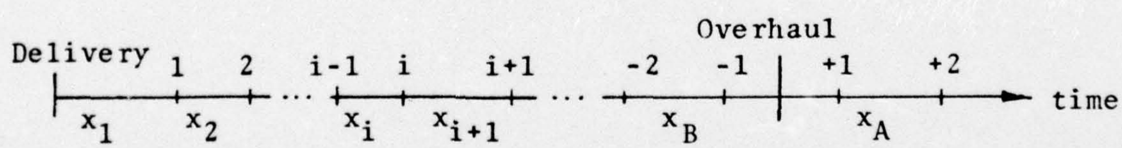
3.3 Continuous Bivariate Distributions Applied to Aircraft Failure Data

In this section we describe how bivariate probability density functions associated with certain aircraft failure data could be useful in managing aircraft operations and maintenance programs. Although the discussion is centered on using bivari-

ate gamma distributions the techniques are applicable to other bivariate densities. First we describe the basic random variable of interest and then show how meaningful bivariate models can be constructed.

For an aircraft we take as a r.v. X the operating time between failures on any component for which the univariate gamma distribution is expected to adequately describe the observed failure times. An examination of historical data would indicate candidate components. (We use the term component to represent an individual part or a subsystem of parts.) For applications to be defined later we also require on each aircraft the delivery date and dates of overhauls. These overhauls are on the entire aircraft and not just on the component itself.

On the following time axis we illustrate, for a particular aircraft and component, the random variable and events described above.



Failures occur at epochs 1,2,..., and the observed operating times are $x_1, x_2, \ldots$. Here the observation x_i is taken to be a realization of X_i which is the random variable representing the operating time between the (i-1)st and ith failures. An

aircraft overhaul occurs at the indicated point and the previous failures are labeled -1 and -2; the intervening op-erating time is denoted by x_B. In a like manner the nex failures after overhaul are labeled +1 and +2 and the in vening operating time is x_A. Next we describe some biva models and possible uses.

For some components it is reasonable to expect that operating times X_i and X_{i+1} are related (particularly fo ponents that are repaired and replaced on the same aircr and so we assume the existence of a bivariate distributi describe this dependence. Given bivariate data $(x_i, x_{i+1}$ n=1,2,...,N, we could determine if the sample values x_i x_{i+1} are correlated and if, in fact, the observed data c be fitted adequately with any of the existing bivariate distributions. To describe some possible uses of bivari probability models in this context we assume the data co be fitted adequately.

If the random variables X_i and X_{i+1} can be describe a bivariate distribution and if, for a particular aircra an observation x_i is given, then the conditional distrib of $X_{i+1}|x_i$ could be used to predict the time of the next on this aircraft, at least due to the component being co The regression function, that is, the mean of the condit distribution could be used to predict the mean time of t failure. This would be an improvement over a prediction

equal to the mean time between failure (MTBF) which is typically used and is computed as if the random variables X_i and X_{i+1} are independent and identically distributed.

Aircraft components generally are assumed to exhibit failure rates which vary with time according to the classic bath tub curve. Thus, failures occur more rapidly at first then decrease to a somewhat constant level and finally increase in frequency. Failures in these three periods commonly are called initial, chance, and wear-out failures, respectively (Mann, et. al. (1974)). A comparison of the bivariate densities associated with the sample operating times on either side of several failures could suggest a pattern of component aging.

One of the prime considerations in simulation studies involving a composite of aircraft operations and support functions is in generating realistic component failures. Certainly, if times to failure are dependent then a mechanism which allows for pairs of observations with the proper correlation would be an improvement over a model which ignored the dependence.

A major objective of aircraft overhaul is to restore the aircraft to a more reliable condition. As before we are interested in determining if the overhaul does improve performance. If we assume the existence of a bivariate model to describe the dependence between X_i and X_{i+1} where no overhaul is involved, then we can investigate the effect of overhaul by comparing a

sample distribution associated with (X_i, X_{i+1}) to the distribution which results by fitting the observed data (x_B, x_A) from a random sample of aircraft with an overhaul.

These are but a few of the bivariate r.v.'s and applications which could be described for continuous data.

3.4 Summary

In this part we applied discrete bivariate probability distributions to actual aircraft operations and maintenance data and showed how these distributions could be used in problems related to inventory control and aircraft overhaul. Additionally, we described how continuous bivariate distributions could be used to analyze certain continuous failure data. All of the applications presented were for self-pairing type situations; without doubt, bivariate applications exist also in situations involving a dependency between two separate items. The latter is a more traditional approach for continuous bivariate r.v.'s.

We presented several new data sets related to demand for aircraft spare parts and materiel failure induced aborts. For the demand and abort data we investigated the marginal and joint data and showed how the univariate and bivariate negative binomial distributions fitted the observations. The P values associated with the χ^2 goodness-of-fit test ranged between 0.01 and 0.85 with the average being about 0.36. One of these data

sets had a negative sample correlation coefficient which emphasizes the need for bnb distributions that can describe such dependencies (in Part 2 we discussed such a distribution).

As a possible way to investigate the effect of a noted event on an item's performance we illustrated, with aircraft abort data and the overhaul event, how bivariate distributions could be used. Our methodology involved comparing two bivariate distributions, one defined for r.v.'s on either side of the event and the other defined for a similar r.v. not separated by the event of interest. We used the sample regression functions to compare the distributions.

Extensions to multivariate (beyond two) settings are obvious. Certainly the difficulty of obtaining parameter estimates is compounded though.

In the next part we apply a bivariate exponential distribution to some correlated queueing systems.

PART 4

CORRELATED QUEUEING SYSTEMS

4.1 Introduction and Historical Review

Using simulation techniques Paulson and Beswick (1973) showed the effect of dependent exponential service times on queues in series. In this part we review their work and present a set of recursive formulae useful in simulating the queueing process. We use spectral analytic techniques to show that the effect is indeed statistically significant. Also, we investigate the effect of correlated interarrival and service processes on single server, single stage queues.

First, we describe two physical settings where tandem queues with dependent service times can be expected to arise.

In a paper mill, large rolls of paper typically pass through an inspection or winding operation prior to being cut into smaller rolls. A poor quality roll takes a relatively longer time in the inspection process because defective sections must be removed and splices made. When this same roll reaches the final cutting stage it must be processed more slowly to avoid breaking the splices and to repair them when they do break. Hence process times at the two stages tend to be correlated; indeed, it is conceptually possible that they be highly correlated. The process times at the two stages on any two different rolls would generally be independently distributed. In the current context considerable interest would be centered on

the effect, if any, produced by non-independence of process times at different stages.

Jackson (1954) in discussing queueing systems with phase type service pointed out that a typical sequence of events in the overhaul of an aircraft engine consist of stripping, detailed examination, repairs, assembly and testing. Generally, an engine with a large number of maintenance requirements can be expected to spend more time in each of the latter four phases and so the possible effect of correlated service times on throughput time would be of interest. It is not difficult to envision a host of other situations involving queues in series in which the service times at the various stages for a given customer are correlated.

A large proportion of the literature concerning tandem queues has centered on Poisson arrival processes, exponential service times, and steady state solutions. The assumption of independence of service times is intricately interwoven into the fabric of the traditional birth-death equation approach to finding a transient and steady state solution to the tandem queueing phenomenon. We shall remain within this same framework with the exception that we shall drop the heretofore universal (but tacit!) assumption of mutual independence of all exponential service times. An obvious approach is to use a multivariate exponential distribution with non-zero correlations in place of the usual independent exponential service times. In

our situation it is not clear that the birth-death equation approach can be modified to incorporate dependent service times. Moreover, any such formulation would very likely be analytically intractable. The problem is however amenable to a simulation approach and it is in this way that we assess the effect of departures from independence of service times on steady state system performance.

The bulk of this part is concerned with two stage queues in series since our main concern is showing that a substantial effect on system performance is indeed induced by correlated service times.

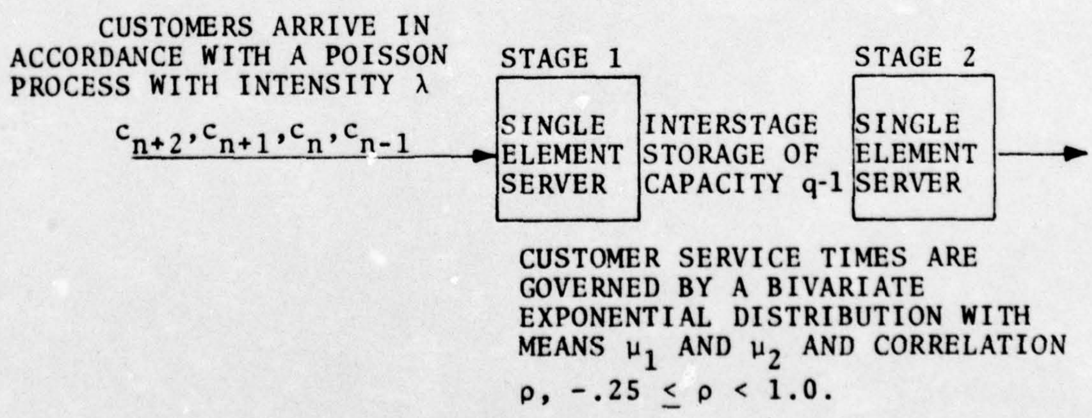
For the unusual single server, single stage queue where the length of a customer's service time is determined, with probability one, by the length of the interarrival interval separating himself and his predecessor, Conolly (1968) gives the waiting time distribution and its moments. It is also shown that this pattern of server behavior results in a drastic reduction of the mean and variance of the waiting time as compared with the conventional M/M/1 system. Conolly refers to this type of system as self-regulating and other results are given by he and Hadidi (1969, 1974). We study, via simulation, this type of system where the dependence between the service time and interarrival interval is assumed to be probabilistic according to a bivariate exponential distribution.

Other investigations involving dependency between the service and interarrival processes have typically allowed for a dependence between a service time and an arrival one interval later than the one we consider; See, for example Lindley (1952) and John (1963). Next we discuss tandem queues.

4.2 The Effect of Correlated Exponential Service Times on Single Server Tandem Queues

4.2.1 The Tandem Queueing System and Recursive Formulae

Consider the tandem queueing process depicted in the following diagram. Customers from an infinite population arrive at a two stage system according to a Poisson process with mean rate λ which we shall, without loss of generality, take to be unity. An unlimited queue is always allowed before the first stage but before the second stage the queue length may be either restricted or unlimited. A single server is allowed at each stage; the service discipline is first-come, first-served.

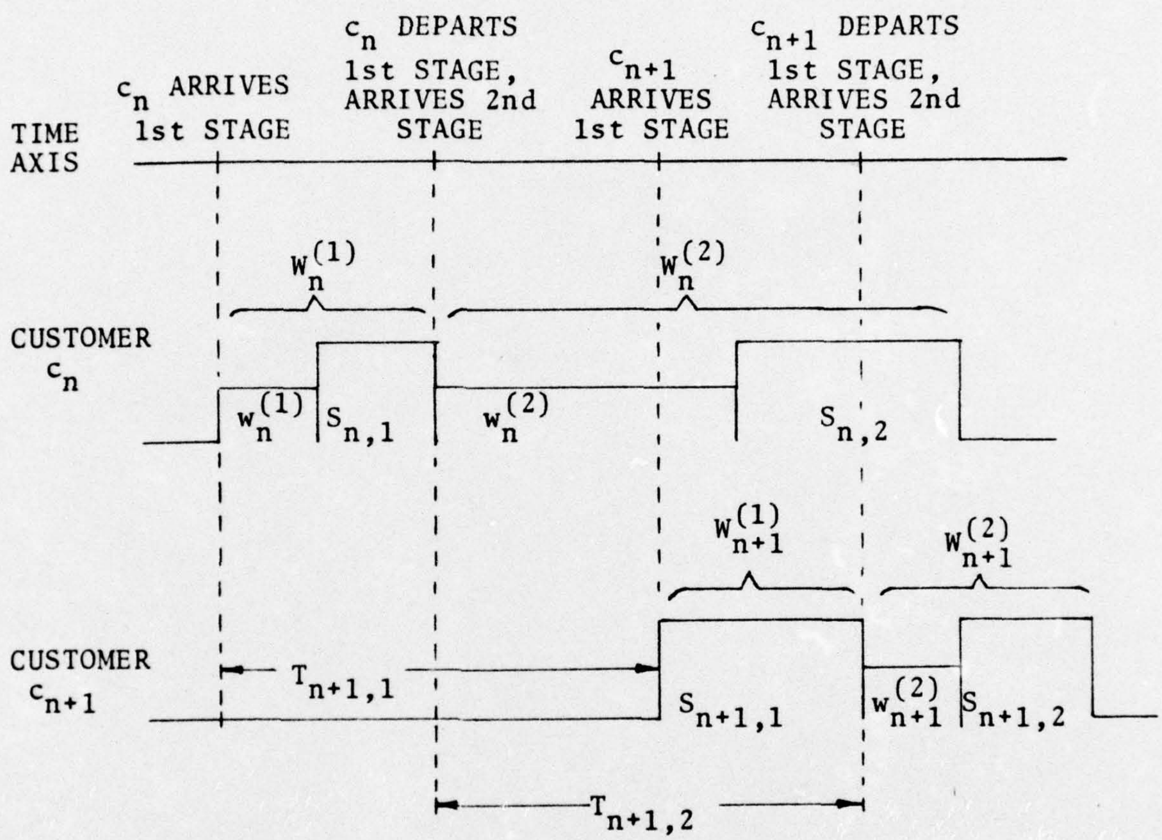


The system performance measure is taken to be mean waiting time per customer and in this section we develop a set of formulae to recursively compute the waiting time per customer. We use the recursive formulae for the unlimited interstage storage case in order to demonstrate precisely how the queueing system consisting of two stages in series with dependent service times is related to a single server system with interdependent arrival and service processes as discussed by Bhat (1969). An interpretation by Conolly (1968) for a special type of this latter interdependence is shown to be helpful in suggesting why mean waiting time is affected by correlated service times.

Denote by $(T_{n,1}, T_{n,2})$ the times between arrival epochs of customers c_{n-1} and c_n at the first and second stages and let c_n experience the service times $(S_{n,1}, S_{n,2})$ at each stage, n=1,2,... . The sequences of interarrival times $(T_{n,1}, T_{n,2})$ and the $(S_{n,1}, S_{n,2})$ for different customers are both assumed to be mutually independent and independent of each other.

We take $(w_n^{(1)}, w_n^{(2)})$ to be the waiting times, excluding service, and $(W_n^{(1)}, W_n^{(2)})$ to be the total waiting times, of customer c_n, at the respective stages, n=1,2,... . We illustrate these definitions with an arbitrary combination of arrival and service times in the following diagram. The illustration is for two queues in series with unlimited

interstage storage; diagrams like this are useful in developing the recursive formulae for the different cases to be presented.



Case A. <u>Two stage queues in series, unlimited interstage storage</u>

Customer c_{n+1}'s total waiting time at the first stage and interarrival time at the second stage are given by

$$W_{n+1}^{(1)} = \begin{cases} S_{n+1,1}, & \text{if } T_{n+1,1} \geq W_n^{(1)} \\ W_n^{(1)} - T_{n+1,1} + S_{n+1,1}, & \text{if } T_{n+1,1} < W_n^{(1)} \end{cases} \tag{4.1}$$

and

$$T_{n+1,2} = \begin{cases} T_{n+1,1} - W_n^{(1)} + S_{n+1,1}, & \text{if } T_{n+1,1} \geq W_n^{(1)} \\ S_{n+1,1}, & \text{if } T_{n+1,1} < W_n^{(1)}. \end{cases} \tag{4.2}$$

The condition in (4.1) and (4.2) that $T_{n+1,1} \geq W_n^{(1)}$ simply means that c_{n+1} arrives at server one after c_n has departed, and likewise $T_{n+1,1} < W_n^{(1)}$ means c_{n+1} arrives before c_n leaves.

Similar to (4.1), c_{n+1}'s waiting time at the second stage is

$$W_{n+1}^{(2)} = \begin{cases} S_{n+1,2}, & \text{if } T_{n+1,2} \geq W_n^{(2)} \\ W_n^{(2)} - T_{n+1,2} + S_{n+1,2}, & \text{if } T_{n+1,2} < W_n^{(2)}. \end{cases} \tag{4.3}$$

The above diagram illustrates (4.1), (4.2), and (4.3) for $T_{n+1,1} \geq W_n^{(1)}$ and $T_{n+1,2} < W_n^{(2)}$. Similar diagrams result for the remaining conditions.

In an obvious way, we can use these relationships to build up a set of recursive formulae for any number of stages in series where the interstage storage between stages is unlimited (See Appendix B).

Since each customer must proceed through both stages the output of the first stage becomes the input of the second stage and therefore we have, in steady state, that the time interval between arrivals at the second stage satisfies a Poisson process with the same interarrival intensity parameter λ as the input distribution (Burke (1956)). Unlike the first stage, however, c_n's service time at the second stage is correlated with the interarrival time there. In the above diagram, this corresponds to a correlation between $S_{n+1,2}$ and $T_{n+1,2}$. This result is apparent from (4.2) since $S_{n+1,1}$ and $S_{n+1,2}$ are dependent by assumption.

If $S_{n+1,1}$ and $S_{n+1,2}$ are independent as is usually assumed for two stage series systems then each stage, in steady state, can be analyzed independently and since $T_{n+1,2}$ and $S_{n+1,2}$ are independent, as are $T_{n+1,1}$ and $S_{n+1,1}$, the regular M/M/1 results obtain for each stage.

Bhat (1969) describes five different classes of single server first-come first-served, systems with Poisson input and exponential service times which result from relaxing some of the assumptions of independence which are typically assumed. These classes represent more realistic operating systems than those with assumptions of independence; Bhat further points out that more work needs

to be done on these problems than the limited amount reported at that time. One of these classes is for systems with interdependent arrival and service processes as is the case here for $T_{n+1,2}$ and $S_{n+1,2}$.

Conolly (1968) and Conolly and Hadidi (1969, 1974) have studied a dependent structure somewhat similar to this wherein the ratio of service time to interarrival time is constant for all n; they give transient as well as steady state results for the system. Conolly showed numerically that this pattern of server behavior results in a drastic reduction in the mean and variance of the waiting time as compared with a conventional M/M/1 queue. It was noted by Conolly that this kind of server behavior is to be expected from a well regulated service facility where the server adjusts the service time of a customer according to that customer's interarrival time, which the server observes without error. In this way, a long interval gives rise to a long service time, and short intervals corresponding to a succession of rapid arrivals, are followed by correspondingly short service times. This regulated behavior therefore prevents a long queue from forming and cuts down on the mean and variance of the waiting time in the system.

Returning to the two stages in series problem under study we see that this system, via equation (4.2), can be viewed as a type of self-regulated system since

$S_{n+1,2}$ and $T_{n+1,2}$ are related, although not in the deterministic way assumed by Conolly. It will be demonstrated later that our type of stochastic dependence between $S_{n+1,2}$ and $T_{n+1,2}$ gives rise to results which are consistent with Conolly's. This artificial way of viewing the system as a self-regulating device is employed solely to make the effects seem more reasonable and in no way influence the results.

Case B. <u>Two stage queues in series, no interstage storage</u>.

For this case, c_n's total waiting time at the second stage, $W_n^{(2)}$, is always equal to the $S_{n,2}$ so the only quantity of interest here is $W^{(1)}$. Since there is restricted (zero) interstage storage, the phenomenon of blocking occurs and so the waiting time computation is a bit more complicated than in Case A.

Blocking of stage one occurs when a customer who has been served there is denied entry into the second stage because no space remains in the queueing area for stage two. The customer therefore stays in the first stage and prevents that server from accepting a waiting customer for service. In effect, the first server's utilization is diminished (Saaty (1961)).

The total waiting time for c_{n+1} at the first server is given by one of four relationships depending upon the algebraic sign of $W_n^{(1)} - T_{n+1,1}$ -- that is, upon

whether or not c_{n+1} arrives at stage one before o

c_n leaves.

For $T_{n+1,1} < W_n^{(1)}$,

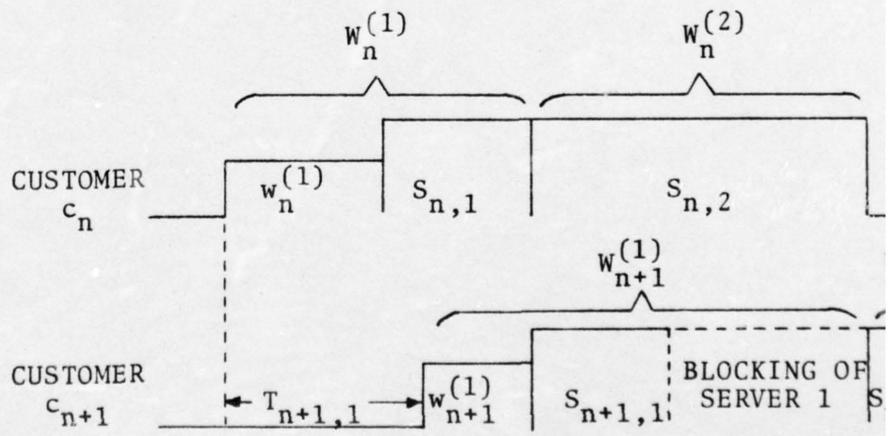
$$W_{n+1}^{(1)} = \begin{cases} W_n^{(1)} - T_{n+1,1} + S_{n+1,1}, & \text{if } S_{n+1,1} \geq S_{n,2} \\ W_n^{(1)} - T_{n+1,1} + S_{n,2}, & \text{if } S_{n+1,1} < S_{n,2} \end{cases}$$

and for $T_{n+1,1} \geq W_n^{(1)}$,

$$W_{n+1}^{(1)} = \begin{cases} S_{n+1,1}, & \text{if } S_{n+1,1} \geq W_n^{(1)} - T_{n+} \\ W_n^{(1)} - T_{n+1,1} + S_{n,2}, & \text{if } S_{n+1,1} < W_n^{(1)} - T_{n+} \end{cases}$$

The following diagram illustrates (4.4)

$S_{n+1,1} < S_{n,2}$.



Similarly, the other conditions can be verified.

Case C. Two stage queues in series, interstage storage capacity equals one.

As in the previous case blocking can occur at the first stage but here a customer's total waiting time at stage two can exceed the service time since interstage storage is permitted on a restricted basis.

The total waiting time of c_{n+1} at server one depends on the algebraic sign of $W_n^{(1)} - T_{n+1,1}$.

If $T_{n+1,1} < W_n^{(1)}$,

$$W_{n+1}^{(1)} = \begin{cases} W_n^{(1)} - T_{n+1,1} + S_{n+1,1}, & \text{if } S_{n+1,1} \geq W_n^{(2)} - S_{n,2} \\ W_n^{(1)} - T_{n+1,1} + W_n^{(2)} - S_{n,2}, & \text{if } S_{n+1,1} < W_n^{(2)} - S_{n,2} \end{cases} \tag{4.6}$$

and c_{n+1}'s interarrival time at server two is

$$T_{n+1,2} = \begin{cases} S_{n+1,1}, & \text{if } S_{n+1,1} \geq W_n^{(2)} - S_{n,2} \\ W_n^{(2)} - S_{n,2}, & \text{if } S_{n+1,1} < W_n^{(2)} - S_{n,2}. \end{cases} \tag{4.7}$$

If $T_{n+1,1} \geq W_n^{(1)}$,

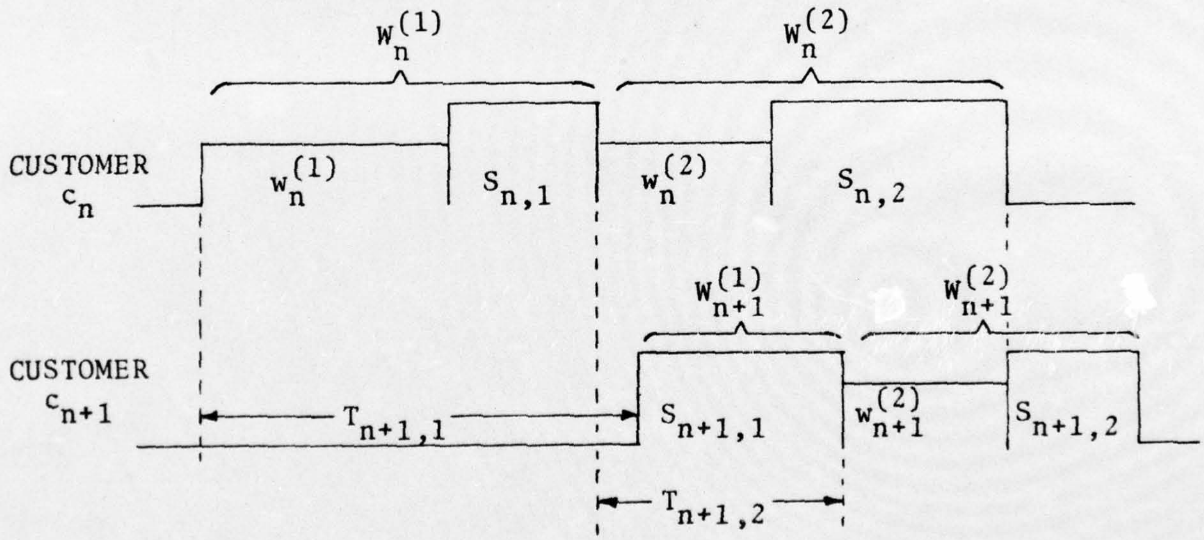
$$W_{n+1}^{(1)} = \begin{cases} S_{n+1,1}, & \text{if } S_{n+1,1} \geq W_n^{(1)} - T_{n+1,1} + W_n^{(2)} - S_{n,2} \\ W_n^{(1)} - T_{n+1,1} + W_n^{(2)} - S_{n,2}, & \text{if } S_{n+1,1} < W_n^{(1)} - T_{n+1,1} + W_n^{(2)} - S_{n,2} \end{cases} \tag{4.8}$$

and

$$T_{n+1,2}=\begin{cases} T_{n+1,1}-W_n^{(1)}+S_{n+1,1}, & \text{if } S_{n+1,1}\geq W_n^{(1)}-T_{n+1,1} \\ & \quad +W_n^{(2)}-S_{n,2} \\ W_n^{(2)}-S_{n,2}, & \text{if } S_{n+1,1}<W_n^{(1)}-T_{n+1,1} \\ & \quad +W_n^{(2)}-S_{n,2}. \end{cases} \tag{4.9}$$

Next the total waiting time of c_{n+1} at stage two, $W_{n+1}^{(2)}$, is computed by using (4.3) in Case A with $T_{n+1,2}$ as defined in (4.7) or (4.9).

The following diagram is descriptive of (4.8) and (4.9) where $T_{n+1,1}\geq W_n^{(1)}$ and $S_{n+1,1}\geq W_n^{(1)}-T_{n+1,1}+W_n^{(2)}-S_{n,2}$.



4.2.2 A Bivariate Exponential Distribution

There are a number of bivariate exponential distributions which could be used to describe the dependence

assumed between S_1 and S_2 (we drop the subscripts n). We choose to use a special case of the bivariate gamma distribution discussed by Wicksell (1933) and Kibble (1941), and more generally by Krishnamoorthy and Parthasarty (1951) and Paulson (1973). The functional form can be written as

$$f(s_1,s_2) = \frac{a}{\theta_1\theta_2} e^{-s_1/\theta_1 - s_2/\theta_2} I_0(2(\frac{ds_1s_2}{\theta_1\theta_2})^{\frac{1}{2}}) \qquad (4.10)$$

where $s_1 \geq 0$, $s_2 \geq 0$, and $I_0(z) = \sum_{k=0}^{\infty} \frac{(z/2)^{2k}}{k!k!}$ is the modified Bessell function of the first kind and order zero. Here $a > 0$, $d \geq 0$, and $a + d = 1$.

The density (4.10) has mean vector

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \theta_1/a \\ \theta_2/a \end{bmatrix}, \qquad (4.11)$$

covariance matrix

$$\Sigma = (\sigma_{ij}) = \begin{bmatrix} \mu_1^2 & d\mu_1\mu_2 \\ d\mu_1\mu_2 & \mu_2^2 \end{bmatrix} \qquad (4.12)$$

and correlation $\rho=d$ with $0 \leq \rho < 1$. The marginal distributions of S_1 and S_2 are exponential with means μ_1 and μ_2 respectively. A generalization of (4.10) due to Paulson admits of correlation values $-.25 \leq \rho < 1$.

The variates (S_1,S_2), given by (4.10), can be readily simulated. Let $\{X_{ij},\ j=1,2,\ldots\}$ be independent identically distributed exponential with mean θ_i for $i=1,2$ and let N denote a geometric random variable with density function

$$\Pr\,[N=i] = p^{i-1}\,q,\ i=1,2,\ldots$$

where p=d and q=a.

Then

$$(S_1,S_2) = \left(\sum_{j=1}^{N} X_{1j},\ \sum_{j=1}^{N} X_{2j}\right) \tag{4.13}$$

has the bivariate exponential distribution (4.10) (see Downton (1970)). The simulation proceeds by simulating N and then adding that number of i.i.d. exponentials according to (4.13).

It would be easy to construct situations in which the service times are negatively correlated. In order that we may readily consider this case we require a bivariate exponential distribution which is conveniently simulated. Such a distribution, which includes the distribution in (4.10) as a special case is due to Paulson (1973). The actual simulation of variates $\underline{S} = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix}$ from that distribution is effected through

$$\underline{S}_\infty = \underline{X}_1 + V_1\,\underline{X}_2 + V_1\,V_2\,\underline{X}_3 + \cdots\ ; \tag{4.14}$$

here $\underline{X}_j$ is a 2-vector of independent exponential variates with mean vector $\begin{bmatrix}\theta_1\\\theta_2\end{bmatrix}$ and the V_j are random 2x2 matrices which take on values in the set { $\begin{bmatrix}0&0\\0&0\end{bmatrix}$, $\begin{bmatrix}1&0\\0&0\end{bmatrix}$ $\begin{bmatrix}0&0\\0&1\end{bmatrix}$, $\begin{bmatrix}1&0\\0&1\end{bmatrix}$ } with probabilities a,b,c,d respectively. All the $\underline{X}_j$'s and V_j's are mutually independent. Note that eventually the product ΠV_j will result in a matrix of zeros and so with probability one $\underline{S}_\infty$ is represented by a finite sum (Kesten (1973), Kohberger (1975)). It turns out that (4.13) is really a special case of (4.14).

The bivariate random variable $\underline{S}_\infty$ in (4.14) has mean vector

$$\underline{\mu}^* = \begin{bmatrix}\mu_1^*\\ \\ \mu_2^*\end{bmatrix} = \begin{bmatrix}\theta_1/(a+c)\\ \\ \theta_2/(a+b)\end{bmatrix} \tag{4.15}$$

and covariance matrix

$$\Sigma^* = \begin{bmatrix}(\mu_1^*)^2 & \frac{ad-bc}{1-d}\theta_1\theta_2\\ \frac{ad-bc}{1-d}\theta_1\theta_2 & (\mu_2^*)^2\end{bmatrix} \tag{4.16}$$

4.2.3 Simulation Results and Interpretation

Simulated results are presented in this section for three cases of interstage storage capacity: (A) infinite ($q=\infty$), (B) zero ($q=1$), and (C) one ($q=2$). For the infinite interstage storage case results will be given for two stages in series for various values of correlation and for two through twenty-five stages in

series for correlation equal unity. The latter depicts how adding stages might affect system performance given correlation $\rho>0$; more precisely, it provides an envelope within which system performance will vary since for a fixed number of stages and utilization correlation unity provides an extremum and correlation zero provides another. In each of the three cases we allow infinite storage before stage one.

In the ordinary case in which the correlation between paired service times is zero a few steady state results are available for comparison purposes.

We have taken the mean arrival rate to be unity and so the steady state utilization, ν, at stage i is simply the mean service time μ_i. It will suffice for our purposes to take $\mu_1 = \mu_2 = \mu$ since similar steady state behavior will obtain for $\mu_1 \neq \mu_2$. Furthermore, there do not seem to be many results available for purposes of comparison for Cases B and C when $\mu_1 \neq \mu_2$. For $\lambda = 1$, our system performance measure of mean waiting time (queueing plus service), is equivalent to the expected number in the system.

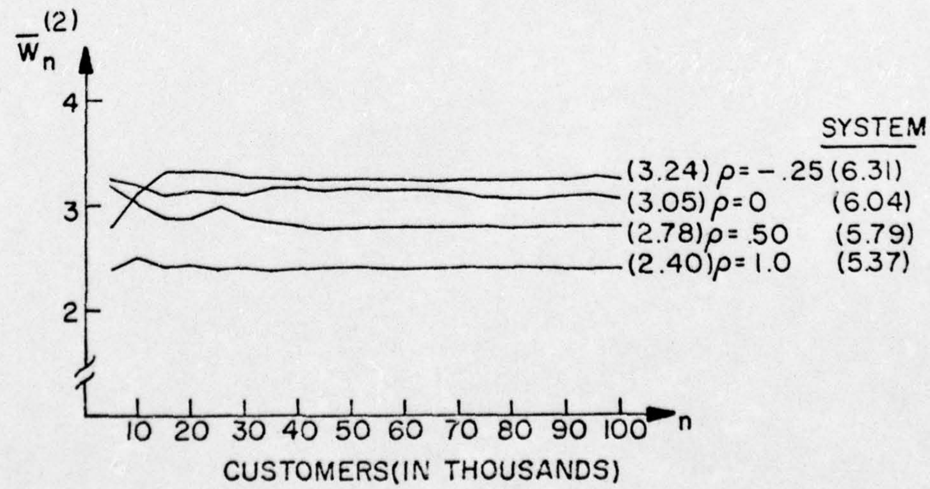
Case A. <u>k stage queues in series, infinite interstage storage</u>.

(Graphs are labeled k Q for k-queues).

Steady state results for k stages in series with no correlation between pairs of service times are avail-

able (Saaty(1961)) and we have that the expected number of customers at each stage is $\nu/(1-\nu)$ and $k\nu/(1-\nu)$ in the system.

The following diagram provides for two stages in series the mean waiting time at the second stage, $\overline{W}_n^{(2)}$, for $\nu = 0.75$ and $\rho = -0.25$, 0, 0.50, and 1.0. In this case the mean waiting time at the first stage is independent of ρ since no blocking occurs and hence it suffices to examine the mean waiting time at the second stage to determine the effects of correlated service times. In some of the simulation results to follow we replicate, many times, runs of much shorter length; here we choose to illustrate the mean waiting time as a function of n with one very long run. Long runs, such as



this one, may be considered as being composed of replications of smaller runs where the starting condition of a new replication is the ending value of the previous replication (Conway(1963)). From the diagram it is clear that each graph tends to stabilize for increasing n in accordance with the law of large numbers.

The numbers adjacent to the values of ρ in the diagram are the mean waiting times for the second stage and mean waiting times in the system after 100,000 service completions. We point out that for $\rho=0$ the mean waiting times at stages one and two are 2.99 and 3.05 respectively and these are in close agreement with the expectation of 3.0 for this utilization. For $\rho\neq0$ we see here a bonus attached to positive correlation in service times since system performance improves with increasing correlation. On the other hand, system performance deteriorates with negative correlation.

Each illustration like this one has a starting condition based on the mean waiting time from a pilot run and then we omit the waiting times of the first 1,000 customers in the actual computations shown.

Figure 3 gives system performance for different values of ν and for $\rho=0,1$. These graphs are intended to show that there is no discernable effect due to correlation $\rho>0$ at utilization $\nu = 0.6$ but as ν increases from

0.6 to 0.9 a definite trend appears.

Figure 4 shows the ratio of mean time in the system for various values of ρ to the mean waiting time in the system at $\rho=0$ (Paulson and Beswick(1973)). These kinds of graphs are based on an average of 100 replications of 1,400 service completions (after an initialization of 400 service completions were discarded).

The solid lines in Figures 4, 7, and 9 depict a smoothed fit to the actual data. Sampling variation, of course, precludes the possibility of obtaining such a smooth fit without extremely long runs or extensive replications but each curve was spot-checked to ascertain whether or not the fit was spurious. In no case was any substantial deviation recorded.

Now we show how these effects are consistent with Conolly's (1968) results for the case of $\nu=0.9$ and correlation of $\rho=1.0$ between the service times in the two stages. Conolly showed for his single server queueing system where the ratio of service time to the interarrival time was constant for all n, that for a utilization of 0.9 (the ratio) the mean waiting time (queueing plus service) was 2.71. For service time independent of interarrival time the steady state expectation, for this utilization, is 9.0. The interarrival time and service time in Conolly's system are perfectly correlated whereas in our system the

two service times are perfectly correlated. It is clear from equation (4.2) that the correlation between the interarrival time at the second stage and the service time there is less than one and so the improvement in system performance for our system should be less than Conolly's (an elementary derivation shows the correlation to be $\nu\rho$ or 0.9 in this case). We see from Figure 3 that the mean waiting time at the second stage after 100,000 customers is 4.13 and indeed the improvement is less.

In Figure 5a we show the mean waiting time as a function of n for five stages in series where the service times are equal at each stage. The graphs are labeled $\bar{W}^{(k)}$ corresponding to the mean waiting time at stage k, $k=1,2,\ldots,5$. We see that the mean waiting time $\bar{W}^{(2)}$, for the second stage, is consistent with the results in Figure 3. The results for $\bar{W}^{(3)}$, $\bar{W}^{(4)}$, and $\bar{W}^{(5)}$ suggest that further improvements in system performance occur over the $\rho=0$ case but the effect seems to approach a limit. The number in parenthesis to the right of $\bar{W}^{(k)}$ is the ratio $\sum_{i=1}^{k} \bar{W}^{(i)}/(k\nu/(1-\nu))$, $2 \leq k \leq 5$. Figure 5b shows this ratio for two through twenty-five stages in series for $\rho=1$. These results were obtained by extending recursive formulae (1), (2), and (3). In this extreme case of correlation, adding stages has an effect on system performance which depends markedly on the utilization rate; e.g., for $\nu=0.7$ system

performance is improved through the first four stages and then is reduced. A utilization of 0.9 gives rise to much improved system performance through twenty-five stages.

Cases B and C. Two stage queues in series, finite (including zero) interstage storage.

For these cases the utilization is effectively reduced in value (Saaty(1961)). The maximum effective utilization is $\nu_{max} = (q+1)/(q+2)$ where the queue in stage two is limited to a length of q-1 units. We consider the cases q=1 and q=2.

Figure 6 shows the mean waiting time at the first stage for q=1 and several values of ν. For this case each customer's waiting time at the second stage is simply the service time there so we are concerned only with the waiting time process at stage one. Figure 7 shows, for stage one, the ratio of mean waiting time at stage one with $\rho \neq 0$ to the mean waiting time at stage one with $\rho = 0$.

Steady state results for the mean number of customers in the system, L, for $\rho=0$, q=1 and with utilization ν are given in Morse (1958); we have that

$$L = 4\nu(2-\nu^2)/((2+\nu)(2-3\nu)). \qquad (4.17)$$

For $\nu=0.4$, 0.5, and 0.6 and for $\rho=0$, the observed (expected) values of L are 1.55 (1.53), 2.87 (2.80), and 7.80 (7.57) respectively. The observed values are from Figure 6.

For $\rho \neq 0$, again we see a dramatic effect in system performance. System performance deteriorates as the correlation ρ increases through positive values and improves as ρ decreases through negative values. Hence when there is no storage allowed before stage two the departure from independence results in significantly different steady state behaviors, especially as the value ν_{max} is approached.

Finally, we consider the case q=2. Figure 8 shows the mean waiting time as a function of n at the first stage and the ending value for the mean waiting time in the second stage. The mean waiting time at stage two was very stable for all values of n so those values will not be illustrated. The effect for $\nu=0.6$ is in the same direction as for q=1 but reverses as ν increases so that for values close to ν_{max} the change in system performance is consistent with the $q=\infty$ case; that is, improvement for $\rho>0$ and deterioration for $\rho<0$. Figure 9 shows the effect in this case for $\rho \neq 0$.

4.2.4 Spectral Analysis of $\{W_n^{(1)}\}$ and $\{W_n^{(2)}\}$

In this section we review briefly the theory of spectral analysis, show the sample power spectra of the time series $\{W_n^{(1)}\}$ and $\{W_n^{(2)}\}$ and finally apply a non-parametric test to the ratio of certain estimated power spectra.

Several authors give complete accounts of the theory and application of spectral analysis; e.g. see Anderson (1971) and Jenkins (1961). Fishman and Kiviat's (1967) paper on the analysis of simulation generated time series is also of direct interest. Our objective here is to review enough of the fundamentals of spectral analysis to motivate the nonparametric test to be presented later.

In general we take $\{Z_t, t \varepsilon T\}$ to be a stochastic process and we let $\{z_t, t \varepsilon T\}$ denote a sequence of observations from the process; the sequence is referred to as a realization of the process or simply as a time series. We take the index set T to consist of discrete, equispaced time points. From the time series we seek to describe the underlying process. Seldom can we determine the form of the multivariate distribution which generated the realization and most often we must make simplifying assumptions even to describe any of the distribution's moments.

We assume that the process is in a particular state of equilibrium where the first and second moments are independent of time. Therefore,

$$E(Z_t) = \mu \tag{4.18}$$

and

$$Cov(Z_t, Z_{t+k}) = \gamma_k \tag{4.19}$$

AD-A031 752 AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OHIO F/G 12/2
ON SOME NEW BIVARIATE NEGATIVE BINOMIAL AND GAMMA DISTRIBUTIONS--ETC(U)
FEB 76 C R MITCHELL

UNCLASSIFIED AFIT-CI-7T-8 NL

2 OF 2
AD
A031752

END

DATE
FILMED
12-76

for all integers k. Given a sample of N observations from the process, we estimate μ and γ_k with $\bar{z}$ and c_k where

$$\bar{z} = \frac{1}{N} \sum_{t=1}^{N} z_t \tag{4.20}$$

and

$$c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}). \tag{4.21}$$

The type of equilibrium described above is called weak or covariance stationarity or stationarity in the wide sense. A study of the time series in terms of its autocovariances (the γ_k) is referred to as a time domain analysis. Another type of analysis is concerned with the frequency content of the time series, namely spectral analysis.

The Fourier cosine transform of the autocovariances $\gamma_0, \gamma_1, \gamma_2, \ldots$, is called the power spectrum. Denoting the power spectrum by $f(\omega)$, we can write

$$f(\omega) = \frac{1}{\pi} [\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos 2\pi\omega k], \quad 0 \leq \omega \leq \tfrac{1}{2} \tag{4.22}$$

and inverting $f(\omega)$ we can express γ_k as

$$\gamma_k = \int_0^{\frac{1}{2}} f(\omega) \cos 2\pi\omega k \, d\omega, \; k=0,1,2,\ldots . \tag{4.23}$$

When k=0 we obtain the variance γ_0 of the process as the integral of the power spectrum:

$$\gamma_0 = \int_0^{\frac{1}{2}} f(\omega)\, d\omega. \tag{4.24}$$

Thus the power spectrum can be considered as a decomposition of the variance at different frequencies.

To get sample results which are statistically consistent we do not estimate the spectrum at a particular frequency but instead estimate the average power about the frequency of concern. The average power corresponds to weighting the autocovariances in the time domain and we typically estimate $f(\omega)$ with the truncated estimate

$$\hat{f}(\omega_j) = \frac{1}{\pi}[\lambda_0\, c_0 + 2 \sum_{k=1}^{m} \lambda_k\, c_k \cos 2\pi\omega_j k] \tag{4.25}$$

where $\omega_j = j/(2m)$, $j=0,1,2,\ldots,m$ and the weights λ_k, $k=0,1,2,\ldots,m$, form a so-called lag window. We choose the Blackman-Tukey "hamming" window,

$$\lambda_k = 0.54 + 0.46 \cos \pi k/m, \; k=0,1,2,\ldots,m. \tag{4.26}$$

In (4.25), the sample autocovariances $c_{m+1}, c_{m+2}, \ldots,$ are omitted since, for m sufficiently large, they should contribute little information. As a result, only m autocovariances need be calculated and savings in computation may be considerable. Considerable care must be used when selecting m, however, because too large a value will increase the variance of the estimates and too small a value will not give enough resolution.

Next we examine several sample power spectra associated with the simulated waiting times for the two-server infinite interstage storage case. We take the simulated values $\{W_n^{(i)}\}$, $n=1,2,\ldots,N$; $i=1,2$, to be time series where, as before, $W_n^{(i)}$ is the total waiting time, queueing plus service, of customer n at server i. Figure 10 shows a portion of the sample spectra for $\{W_n^{(1)}\}$ and $\{W_n^{(2)}\}$, $n=1,2,\ldots,2000$, and for correlation values of $\rho=0$, 0.25, 0.50, and 1.0. Utilization, ν, is 0.90. The 2000 sample values were chosen from the end of a simulation run of length 30,000 to ensure that any possible effects of start-up conditions were eliminated. After making several pilot runs, m in equation (4.25) was set equal to 400. For $\rho=0.50$ and $\rho=1.0$ in Figure 10 it is obvious that the waiting times at the second server give rise to different spectra than the waiting times at the first server.

Since the integral of the power spectrum measures the variance of the process and the area under the sample spectrum should be indicative of the sample variance, we see that the effect of positive correlation is to <u>reduce</u> the variance of the waiting time process. Again this is consistent with Conolly's results (1968) for the single server system in which a customer's service is completely determined by the length of the interarrival interval separating himself and his predecessor. For a utilization

of 0.9, Conolly's system reduces the steady state variance of the waiting time from 81, for the classic M/M/1 system, to 1.16; the sample variance associated with the waiting times at server 2 in Figure 10d for $\rho=1.0$ is 2.05 (the sample variance associated with the waiting times at server 1 is 60.1). Recall from Section 4 that the condition $\rho=1.0$ for the correlation between a customer's service times at the two servers is equivalent to a correlation of ν, or 0.9 in this case, between his interarrival time and service time at the second server. Therefore, the reduction in variance is consistent with Conolly's results since the corresponding correlation in his system is one.

Next we develop a nonparametric test for the hypothesis that $f^{(1)}(\omega) = f^{(2)}(\omega)$, $0 \leq \omega \leq 0.5$, where $f^{(i)}(\omega)$ represents the power spectrum at frequency ω associated with the time series $\{W_n^{(i)}\}$, $n=1,2,\ldots,N$; $i=1,2$. The Blackman-Tukey "hamming" lag window in (4.25) gives rise to spectral estimates which are not independent and so we employ the notion of equivalent independent estimates (Jenkins(1961)) which implies, for this window, that estimates are approximately independent if they are about $5/(4m)$ cycles apart. Since the estimates in (4.25) are separated by a basic frequency of $1/(2m)$ cycles, this spacing of $5/(4m)$ cycles amounts to taking, as independent, those estimates which are separated by an interval of 2.5

times the basic frequency. Since the waiting times are not normally distributed and the assumption of normality is implicit in the development of equivalent independent estimates we take this spacing of 2.5 times the basic frequency simply to be a rough guide. Actually, the normality assumption is more critical for making distributional assumptions about the spectral estimates than for the usage here. To select a practical spacing and to reduce any possible effects of the normality assumption, we take estimates at the frequencies $j/(2m)$, $j=1,4,7,\ldots$, to be approximately independent (the spacing here is 3 times the basic frequency). Therefore, of the 401 estimates in each spectrum partially illustrated in Figure 10, we take 134 estimates at the frequencies $j/800$, $j=1,4,7,\ldots,399$, to be approximately independent.

Now for each approximately independent estimate we can regard the ratio $\hat{f}^{(1)}(\omega)/\hat{f}^{(2)}(\omega)$ as a Bernoulli trial (greater than unity or less than unity) and under the null hypothesis of homogeneity of the two spectra, we can take as a test statistic the number of ratios which are less than unity. Figure 11 shows the ratio for $\rho=0$ and $\rho=0.25$. Of the 134 approximately independent ratios in Figure 11a, 64 are less than unity and in Figure 11b, 43 of the 134 ratios are less than unity. Under the null hypothesis, a ratio greater than unity is as equally likely as a ratio

less than unity and taking a normal approximation to the implied binomial distribution, we have a probability of 0.31 associated with observing 64 or fewer ratios less than unity in 11a and a probability of .001 associated with observing 43 or fewer ratios less than unity in 11b. Although not illustrated, the results for $\rho=0.50$ and $\rho=1.0$ are even more conclusive: for $\rho=0.50$, 23 of the 134 approximately independent estimates are less than unity and for $\rho=1.0$, none of the ratios are less than unity. Therefore, for the case presented here we have good statistical evidence that the power spectra associated with the waiting times at each server are not homogeneous for correlation $\rho>0$. We expect similar results for other values of correlation, utilization, and interstage storage to obtain.

In the next section we study in a similar way single server queues (not in tandem) which have correlated interarrival and service processes.

4.3 Single Server Queues with Correlated Interarrival and Service Processes

In this section we are concerned with a single server, first-come, first-served queueing system where we depart from the usual assumption of independence by taking a customer's interarrival interval and subsequent service times to be correlated according to the bivariate exponential distribution described in 4.2.2. Here the customer's interarrival interval is measured between his arrival time and that of his predecessor. It is assumed

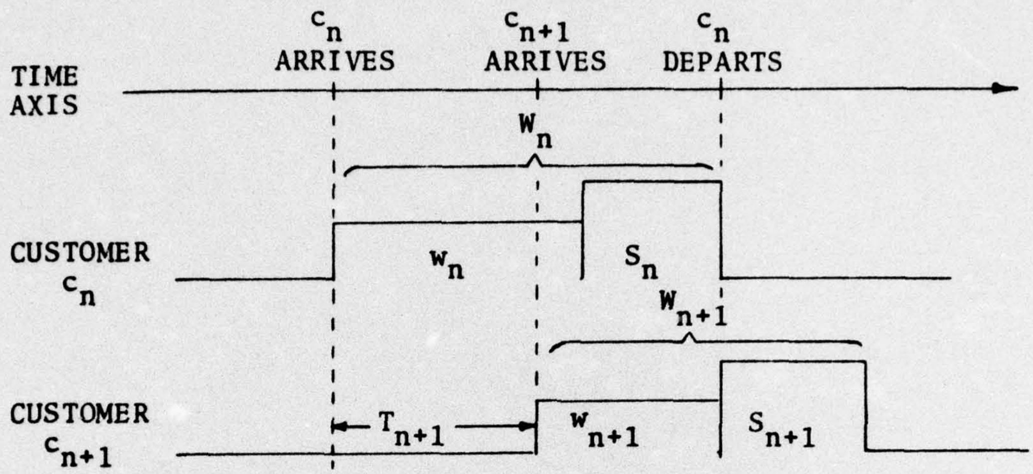
that customers from an infinite population arrive at a single stage according to a Poisson process with rate λ, which we take, without loss of generality, to be unity; an unlimited queue is allowed.

As in the previous section, the system performance measure is taken to be the mean waiting time per customer and in this section we show a formula to recursively compute the waiting time per customer. We give simulated results which show the effect of correlation and apply the nonparametric test of the last section to show that the effect is indeed statistically significant.

Denote by T_n the time between arrival epochs of customers c_{n-1} and c_n to the queue and let c_n experience the service time S_n, $n=1,2,3,\ldots$. The sequence of interarrival times $\{T_n\}$ and service times $\{S_n\}$ for different customers are both assumed to be independent; for customer c_n we assume the r.v. (T_n, S_n) has the bivariate exponential distribution given in (4.10) with $\mu_1=1$ (to be associated with T_n) and $\mu_2=\nu$ (for S_n), $0 < \nu < 1$, so that the steady state utilization is $\mu_2=\nu$.

For customer c_n we define w_n to be the waiting time, excluding service, and W_n to be the total waiting time, $n=1,2,\ldots$; the following diagram illustrates the definitions. It is obvious that a recursive formula for W_{n+1} is

$$W_{n+1} = \begin{cases} W_n - T_{n+1} + S_{n+1}, & \text{if } T_{n+1} < W_n \\ S_{n+1}, & \text{if } T_{n+1} \geq W_n \ . \end{cases} \tag{4.27}$$



For T_{n+1} and S_{n+1} independent as is normally assumed it can be shown (Morse(1958)) that the mean waiting time per customer, in steady state, is $\nu/(1-\nu)$. At the other extreme for $\rho=1$, Conolly (1968) gives the distribution of the waiting time and its mean and variance for the case $S_{n+1} = \nu T_{n+1}$ for all n. Our results are for other values of correlation. Next we use (4.27) to show simulated results.

In Figure 12 we show how nonzero correlation affects the mean waiting time for $\nu=0.70$. (The simulations are performed the same as described in the previous section.) For zero correlation the expected waiting time in steady state is 2.333 and we see that the simulated results are in close agreement (2.323). Conolly shows for his system ($\rho=1$) that the expected waiting time for this value of ν is 1.427 and again the agreement is very good (1.421). For positive correlation we have a benefit in system performance in that mean waiting time decreases; negative correlation degrades the process.

Figure 13 shows the ratio of mean time in the system for various values of ρ to expected waiting time in the system at $\rho=0$ and we see that the effect of nonzero correlation is greatest for large utilizations. In every case simulation results were in close agreement with known results (for $\rho=0$ and $\rho=1$).

Next we examine sample power spectra asoociated with the waiting time process and test the hypothesis $f_0(\omega) = f_\rho(\omega)$, $0 \le \omega \le ½$, where $f_0(\omega)$ is the power spectrum associated with the waiting time process at $\rho=0$ and likewise, $f_\rho(\omega)$ corresponds to $\rho\neq0$.

We take the simulated values $\{W_n\}$, $n=1,2,\ldots,2000$, to be a time series and Figure 14 shows a portion of the sample power spectra for $\rho=0$, 0.50, and 1.0; utilization is 0.70. The spectra appear different and we suspect that the variances of the waiting time processes decrease with positive correlation due to the relation of the illustrated graphs. In fact, the corresponding simulated waiting time series have variances as follows: $\rho=0$, 4.763; $\rho=0.50$, 2.141; and $\rho=1.0$, 0.514. (The expected variance for $\rho=0$ is 5.444 (Morse 1958) and Conolly's model for $\rho=1$ gives rise to a variance of 0.613.) Coupled with the simulated result that $\rho=-0.25$ leads to a variance of 10.451 we see that the effect of positive correlation is to <u>reduce</u> the variance of the waiting time process and negative correlation causes an <u>increase.</u>

As a function of the approximately independent estimates defined in 4.2.4, Figure 15 shows the ratios, $\hat{f}_0(\omega)/\hat{f}_\rho(\omega)$, for $\rho=0.5$ where the caret signifies that we are using sample estimates. Recall that we regard each ratio as a Bernoulli trial (greater than unity or less than unity) and under the null hypothesis of homogeneity of the two spectra we take as a test statistic the number of ratios which are less than unity. In the figure, 22 of the 134 approximately independent ratios are less than unity which is very strong evidence that the null hypothesis is false; for other cases of ρ the number of ratios less than unity are: $\rho=-0.25$, 87; $\rho=0.25$, 43 and $\rho=1.0$, none. Additionally we applied the test to $\hat{f}_{0.25}(\omega)/\hat{f}_{0.50}(\omega)$ and got 38 of the 134 approximately independent ratios less than unity. We reject the implied null hypotheses in all cases and conclude that the waiting time process as a function of ρ leads to different power spectra.

For interest's sake we investigated the system under study for a different bivariate exponential distribution. Primarily due to the ease with which the variates can be simulated, we chose the bivariate exponential distribution of Marshall and Olkin (1967). If the r.v.'s U, V and W are independent exponentials with parameters λ_1, λ_2 and λ_{12}, respectively, then the bivariate r.v. (T, S), where T=min(U, W), S=min (V, W), has the indicated distribution. It can be shown that T and S have means $1/(\lambda_1 + \lambda_{12})$ and $1/(\lambda_2 + \lambda_{12})$, respectively, and the correlation

between the two is $\lambda_{12}/(\lambda_1 + \lambda_2 + \lambda_{12})$. For the proper choice of parameters then, we can generate a bivariate r.v. where $E[T]=1$, $E[S]=\nu$ and the correlation is ρ.

Figure 16 shows comparative results for the Marshall and Olkin model and the Wicksell-Kibble special case in (4.10). Although not illustrated, we applied the homogeneity test to the two sample spectra shown with the result that 65 of 134 approximately independent ratios $(\hat{f}_{W-K}(\omega)/\hat{f}_{M-O}(\omega))$ were less than unity. The probability is approximately 0.37 of observing 65 or fewer ratios less than unity if the hypothesis $f_{W-K}(\omega) = f_{M-O}(\omega)$ is true so we fail to reject it. We expect similar results for other values of correlation and utilization to obtain.

4.4 Summary

We have extended the work of Paulson and Beswick (1973) for the effect of dependent exponential service times on the system performance of tandem queueing systems. We assumed that each queue has a single server and the service discipline is first-come first-served.

For the two stage queueing system we derived recursive formulae for the total waiting time per customer at each queue for the cases of zero, one, and infinite interstage storage. Simulation results for the mean waiting time, under the assumption of correlated service times, showed that the system's behavior is quite sensitive to departures from the traditional

assumption of mutually independent service times, especially at higher utilization rates. For the case of infinite interstage storage, mean waiting time is reduced by positive correlation and increased by negative correlation. This change is reversed, however, for zero interstage storage and depends on the value of the utilization rate for the case where interstage storage equals unity. By using spectral analysis and a nonparametric test applied to the sample power spectra associated with the simulated waiting times we showed that the effect is statistically significant; in addition we showed that the variance of the waiting time process is reduced for positive correlation.

We showed in a precise way how the two stage queueing system with dependent service times and infinite interstage storage is related to a single server system with interdependent arrival and service processes; an interpretation by Conolly (1968) for a special type of this latter interdependence is shown to be useful in suggesting why the mean and variance of the waiting time process are affected by correlated service times.

For correlation equal unity and infinite interstage storage, results were shown for two through twenty-five stages in series; these results provide an envelope within which system performance will vary since for a fixed number of stages and utilization correlation unity provides one extremum and correlation zero provides another.

Additionally, we studied single server, single stage queueing systems wherein a customer's interarrival interval and subsequent service times are governed by a bivariate exponential distribution. For the case of unlimited storage capacity we showed that positive correlation leads to reduced mean waiting times and negative correlation increases the mean waiting times, both more so at higher utilizations. The results were shown to be statistically significant. We also showed that the variance of the waiting time process is reduced for positive correlation and increased for negative correlation. Briefly we investigated the effect of using the Marshall and Olkin (1967) bivariate exponential distribution; results were similar.

PART 5

DISCUSSION AND CONCLUSIONS

Here we have investigated various bivariate non-normal distributions and showed several areas to which they could be applied. Of principal concern were bivariate negative binomial and gamma distributions. In Part 2 we discussed numerous theoretical aspects and fitted the distributions to data, Part 3 illustrated some bivariate approaches for analyzing aircraft operations and maintenance data and lastly, Part 4 showed how bivariate exponential distributions could be applied to queueing systems with certain kinds of correlation. Summary results were given in each part; next, we highlight these results and suggest areas for future research.

One bnb distribution was obtained by convolving the Paulson-Uppuluri (1972) bivariate geometric distribution which is defined by a certain characteristic-functional equation. The distribution has six parameters and admits of positive or negative correlation and linear or nonlinear regression functions. Shown were the moments to order two and for special cases, the regression function, a recursive formula for the cell probabilities, a method of moments parameter estimation technique, the likelihood equations, the differential-difference equations and for maximum likelihood estimation, a necessary relationship for the parameters. Certain analogous properties were shown for a dual bivariate gamma distribution. For both of these distributions

areas for future research include determining an infinitely divisible representation for the full model (all six parameters) and developing parameter estimation techniques. Both areas represent sizeable tasks.

We discussed another bnb distribution which resulted by reducing a certain trivariate negative binomial distribution with independent marginals. Moments to order two and the regression function were provided. This distribution has marginals whose characteristic functions are of the form $[1+\theta(1-e^{it})]^{-\nu_i}$, whereas the aforementioned distribution's marginal characteristic functions are like $[1+\theta_i(1-e^{it})]^{-\nu}, i=1,2$. Although not investigated here we suspect that bnb distributions which allow θ and ν to vary would have practical value.

That these bnb distributions should be useful was illustrated by analyzing sample data sets, some with negative correlation and nonlinear regression.

Another major effort centered on forming bivariate r.v.'s related to particular aircraft operations and maintenance problems. For a random sample we showed that the negative binomial distribution could be used to adequately describe demands for aircraft spare parts for single time periods (univariate) and adjacent time periods (bivariate). An application for fly-away kits was discussed. Additionally, we investigated aircraft abort data for single six month periods and adjacent six month periods and showed how the univariate and bivariate negative

binomial distributions fitted the data. The P values associated with the χ^2 goodness-of-fit test ranged between 0.01 and 0.85 with the average being about 0.36. With these abort data we illustrated a possible way to examine the effect of a noted event on an item's performance. Here we were interested in the effect of overhaul on an aircraft's performance when measured by aborts for six month periods. Basically the method involved comparing two bivariate distributions, one defined for r.v.'s on either side of the event (overhaul) and the other defined for a similar r.v. not separated by the event. We used the regression functions to compare the sample distributions. For our data, aborts increased after overhaul but since our analysis was limited in certain ways we were unable to conclude that the rise was due to overhaul. We intend to investigate this application more in the future.

The last part showed the effect of certain correlated r.v.'s on the system performance of tandem and single stage queueing systems. A bivariate exponential distribution was used. In both cases we assumed that arrivals were according to a Poisson process, the service discipline was first-come, first-served and a single server was available.

For the two stage tandem queueing system we showed, via simulation, that the mean waiting time is quite sensitive to departures from the traditional assumption of mutually independent service times, especially at higher utilizations. For

the case of infinite interstage storage, mean waiting time is reduced by positive correlation and increased by negative correlation. This change is reversed, however, for zero interstage storage and depends on the value of the utilization rate for the case where interstage storage equals unity. A result by Conolly (1968) was shown to be useful in explaining the effect for the infinite interstage storage case.

For single stage queueing systems where a customer's interarrival interval and subsequent service times are correlated we showed that positive correlation reduces mean waiting time and negative correlation increases mean waiting time. The storage area was assumed to be infinite.

By using spectral analysis and a nonparametric test applied to the sample power spectra associated with certain simulated waiting times we showed the effect, in both cases, to be statistically significant.

PART 6

REFERENCES

Abramowitz, M. and Stegun, I.A., eds. (1964). Handbook of Mathematical Functions. Superintendent of Documents, U.S. Government Printing Office, Washington, D.C.

Anderson, T. W. (1971). The Statistical Analysis of Time Series, John Wiley & Sons, Inc., New York.

Arbous, A. G. and Kerrich, J. E. (1951). "Accident Statistics and the Concept of Accident-Proneness," Biometrics, Vol. 7, 340-432.

Arbous, A. G. and Sichel, H. S. (1954). "New Techniques for the Analysis of Absenteeism Data," Biometrics, Vol. 41, 77-90.

Arnold, B. C. (1967). "A Note on Multivariate Distributions with Specified Marginals," J. Amer. Statist. Ass., Vol. 62, 1460-1.

Bates, G. E. and Neyman, J. (1952). "Contributions to the Theory of Accident Proneness, I," University of California Publications in Statistics, 215-54.

Bhat, U. N. (1969). "Queueing Systems with First-Order Dependence," Opsearch, Vol. 6, No. 1, 1-24.

Block, H. W. (1975). "Physical Models Leading to Multivariate Exponential and Negative Binomial Distributions," Rensselaer Polytechnic Institute, O. R. and S. Research Paper No. 37-75-P3.

Boswell, M. T. and Patil, G. P. (1970). "Chance Mechanisms Generating the Negative Binomial Distributions," published in Random Counts in Models and Structures, Vol. 1, edited by G. P. Patil, Penn. State Univ. Press, 3-22.

Burke, P. J. (1956). "The Output of a Queueing System," Operations Research, Vol. 4, 699-704.

Clark, J. R. (1972). Properties of a New Multivariate Geometric and Negative Binomial Distribution with Possible Applications, M.S. thesis, Univ. of Tenn.

Conolly, B. W. (1968). "The Waiting Time Process for a Certain Correlated Queue," Operations Research, Vol. 16, 1006-1015.

Conolly, B. W. and Hadidi, N. (1969). "A Correlated Queue," J. Appl. Prob., Vol. 6, No. 1, 122-136.

Conolly, B. W. and Hadidi, N. (1974). "A Comparison of the Operational Features of Conventional Queues with a Self-Regulating System," Appl. Statist., Vol. 18, 41-53.

Conway, R. W. (1963). "Some Tactical Problems in Digital Simulation," Management Science, Vol. 16, No. 1, 47-61.

Cross, K. (1970). A Gradient Projection Method for Constrained Optimization, Report No. K-1746, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Dade, M. (1973). Examples of Aircraft Scheduled-Maintenance Analysis Problems, The Rand Corp., R-1299-PR.

Downton, F. (1970). "Bivariate Exponential Distributions in Reliability Theory," J. Royal Statist. Society, Series B, Vol. 32, 408-417.

Draper, N. R. and Smith, H. (1966). Applied Regression Analysis, John Wiley & Sons, Inc., New York.

Edwards, C. B. and Gurland, J. (1961). "A Class of Distributions Applicable to Accidents," JASA, Vol. 56, 503-517.

Faucett, W. M. and Gilbert, R. D. (1966). Characteristics of Demand Distributions for Aircraft Spare Parts, Research and Engineering Departments (ERR-FW-512), General Dynamics, Fort Worth Division.

Fishman, G. S. and Kiviat, P. J. (1967). "The Analysis of Simulation-Generated Time Series," Management Science, Vol. 13, 525-557.

Gibbons, J. D. (1971). Nonparametric Statistical Inference, McGraw-Hill Book Company, New York.

Guldberg, A. (1934). "On Discontinuous Frequency Functions of Two Variables," Skand. Aktuar., Vol. 17, 89-117.

Harris, R. (1968). "Reliability Applications of a Bivariate Exponential Distribution," Operations Research, Vol. 16, No. 1, 18-27.

Hawkes, A. G. (1972). "A Bivariate Exponential with Applications to Reliability," J.Royal Statist. Society, Series B, Vol. 34, 129-131.

Holgate, P. (1964). "Estimation for the Bivariate Poisson Distribution," Biometrika, Vol. 51, 241-5.

Jackson, R. R. P. (1954). "Queueing Systems with Phase Type Service," Operations Research, Vol. 5, 109-120.

Jenkins, G. M. (1961). "General Considerations in the Analysis of Spectra," Technometrics, Vol. 3, No. 2, 133-166.

John, F.I. (1963). "Single Server Queues with Dependent Service and Interarrival Intervals," J. Soc. Indust. Appl. Math., Vol. 11, 526-534.

Johnson, N. L. and Kotz, S. (1969). Discrete Distributions, Houghton Mifflin Co., Boston.

Jury, E. I. (1964). Theory and Application of z-Transform Method, John Wiley & Sons, Inc., New York.

Kemp, C. D. (1970). "Accident Proneness and Discrete Distribution Theory," published in Random Counts in Scientific Work, Vol. 2, edited by G. P. Patil, Penn. State Univ. Press, 41-66.

Kendall, M. G. and Stuart, A. (1973). The Advanced Theory of Statistics, Vol. 2, Hafner Publishing Co., New York.

Kesten, H. (1973). "Random Difference Equations and Renewal Theory for Products of Random Matrices," Acta Mathematica, Vol. 131, 207-248.

Kibble, W. F. (1941). "A Two-Variate Gamma Type Distribution," Sankhya, Vol. 5, 137-150.

Kohberger, R. C. (1975). "On Certain Multivariate Exponential Distributions," Ph.D. Thesis, Rensselaer Polytechnic Institute, Troy, New York.

Krishnamoorthy, A. S. and Parthasarty, M. (1951). "A Multivariate Gamma Type Distribution," Ann. Math. Statist., Vol.22, 549-557.

Lindley, D. V. (1952). "The Theory of Queues with a Single Server," Proc. Camb. Phil. Soc., Vol. 48, 277-289.

Lundberg, O. (1940). On Random Processes and their Application to Sickness and Accident Statistics, Uppsala: Almquist and Wiksell.

Mann, N. R., Schafer, R.E., and Singpurwalla, N.D. (1974). Methods for Statistical Analysis of Reliability and Failure Data, John Wiley & Sons, Inc., New York

Mardia, K. V. (1970). Families of Bivariate Distributions, Hafner Publishing Co., New York.

Marshall, A. W. and Olkin, I. (1967). "A Multivariate Exponential Distribution," J.Amer. Statist. Assoc., Vol. 62, 30-44.

Morse, P. M. (1958). Queues, Inventories and Maintenance, John Wiley & Sons, Inc., New York.

Paulson, A.S. (1973). "A Characterization of the Exponential Distribution and a Bivariate Exponential Distribution," Sankhyā, Series A, Vol. 35, 69-78.

Paulson, A. S. and Beswick, C. A. (1973). "The Effect of Dependent Exponential Service Times on Queues in Series," Department of Operations Research and Statistics, Rensselaer Polytechnic Institute Research Report No. 37-73-P2.

Paulson, A. S. and Uppuluri, V. R. (1972). "A Characterization of the Geometric Distribution and a Bivariate Geometric Distribution," Sankhyā, Series A, Vol. 34, 297-300.

Saaty, T. L. (1961). Elements of Queueing Theory with Applications, McGraw-Hill, New York.

Subrahmaniam, K. (1966). "A Test for "Intrinsic Correlation" in the Theory of Accident Proneness," J. Royal Statist. Society, Series B, Vol. 28, 180-189.

Subrahmaniam, Kocherlakata and Subrahmaniam, Kathleen (1973). "On the Estimation of the Parameters in the Bivariate Negative Binomial Distribution," J. Royal Statist. Society, Series B, Vol. 35, 131-146.

Titchmarsh, E. C. (1964). The Theory of Functions, London, Oxford University Press.

Whittaker, E. T. and Watson, G. N. (1965). A Course of Modern Analysis, Cambridge University Press.

Wicksell, S. D. (1933). "On Correlation Functions of Type III," Biometrika, Vol. 25, 121-133.

Youngs, J.W.T., Geisler, M.A. and Brown, B.B. (1955). The Prediction of Demand for Aircraft Spare Parts Using the Method of Conditional Probabilities, The Rand Corp., RM-1413.

APPENDIX A

A NEW MULTIVARIATE NEGATIVE BINOMIAL DISTRIBUTION

In this appendix we show Clark's (1972) derivation of a new multivariate negative binomial distribution. The distribution results by convolving a multivariate geometric distribution.

The multivariate analogue of (2.13) is

$$\phi(T) = \psi(T)E[\phi(TV)] \tag{A.1}$$

where

$$T = (t_1, t_2, \ldots, t_n),$$

$$\psi(T) = \prod_{j=1}^{n} [1+\frac{p_j}{1-p_j}(1-e^{it_j})]^{-1}$$

and V is the set of all n dimensional diagonal matrices of zeroes and ones (2^n matrices in all). Here $\underline{v}\varepsilon V$ assumes a particular value with probability $a_{\underline{v}}$, $\sum_{\underline{v}\varepsilon V} a_{\underline{v}} = 1$ and $a_{\underline{v}} + a_I < 1$, $\underline{v} \neq \underline{0}$, I where $\underline{0}$ is the zero matrix and I is the identity matrix.

Equation (A.1), which can be written as

$$\phi(T) = \psi(T) \sum_{\underline{v}\varepsilon V} a_{\underline{v}}\phi(T\underline{v}), \tag{A.2}$$

defines a multivariate geometric distribution.

It can be shown that

$$\phi(0,\ldots,0,t_m,0,\ldots,0) = [1+\theta_m(1-e^{it_m})]^{-1} \tag{A.3}$$

where

$$\theta_m = \frac{p_m}{1-p_m}\,[1-\sum_{\underline{v}^*} a_{\underline{v}^*}]^{-1}$$

and $\underline{v}^*\varepsilon V$ such that mth diagonal term is one (2^{n-1} matrices in all). Therefore the marginals are geometric and it follows that

$$E(X_m) = \theta_m,$$

$$\text{Var}\ (X_m) = \theta_m(1+\theta_m),\ m=1,2,\ldots,n. \qquad \text{(A.4)}$$

In (A.2) if $a_{\underline{v}} = 0$ for $\underline{v} \neq \underline{0}$, I we have

$$\phi(T) = \psi(T)(a+d\phi(T)) \qquad \text{(A.5)}$$

where a is associated with $a_{\underline{0}}$ and d with a_I. Convoluting as in (2.25) *yields*

$$\phi_\nu(T) = [a\psi(T)]^\nu[1-d\psi(T)]^{-\nu}$$

$$= [a\psi(T)]^\nu[1+\binom{\nu}{1}d\psi(T)+\binom{\nu+1}{2}(d\psi(T))^2+\ldots] \qquad \text{(A.6)}$$

and inverting gives the multivariate negative binomial distribution

$$g_\nu(x_1,x_2,\ldots,x_n) = C[{}_nF_{n-1}(\nu+x_1,\nu+x_2,\ldots,$$
$$\nu+x_n;\nu,\nu,\ldots,\nu;d\prod_{i=1}^{n}(1-p_i))],$$

where

$$C = \prod_{j=1}^{n}\binom{\nu+x_j-1}{x_j}(1-p_j)^\nu(p_j)^{x_j}$$

and ${}_nF_{n-1}(a_1,\ldots,a_n;\ b_1,\ldots,b_{n-1};z) =$

$$1 + \sum_{j=1}^{\infty} \frac{(a_1)_j \cdots (a_n)_j}{(b_1)_j \cdots (b_{n-1})_j} \frac{z^j}{j!}. \qquad \text{(A.7)}$$

Clark concluded by showing that

$$E[X_m] = \nu\theta_m$$

and

$$\text{Var}\ [X_m] = \nu\theta_m(1+\theta_m),\ m=1,2,\ldots,n;$$

he also showed that the distribution is infinitely divisible.

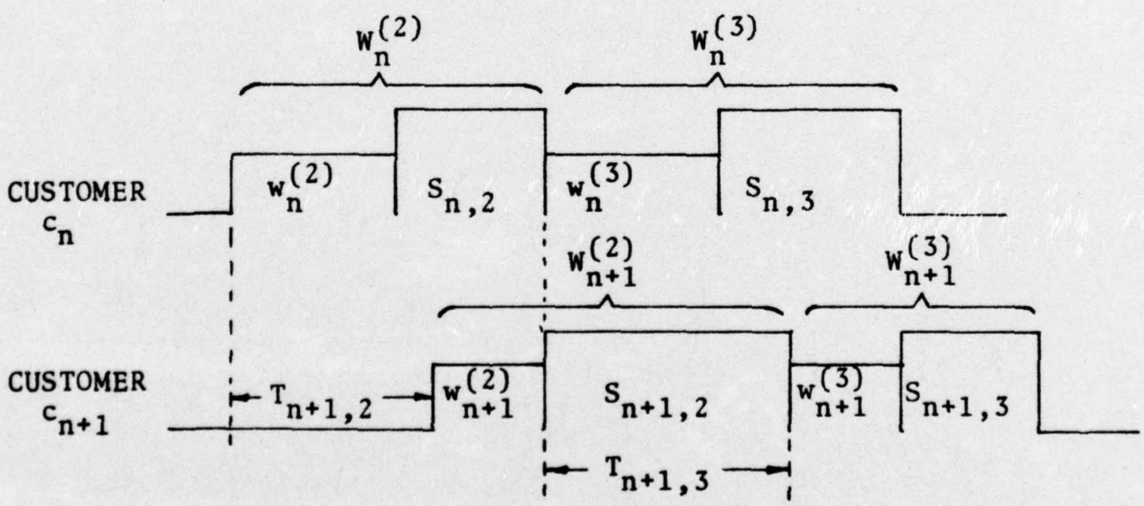
APPENDIX B

RECURSIVE FORMULA FOR WAITING TIMES

In this appendix we show how a set of recursive formulae for the waiting times can be constructed for any number of queues in series where interstage storage is unlimited.

Referring to the following diagram for c_n and c_{n+1}'s queueing and service times at stages two and three (a continuation of the diagram preceding equation (4.1) in the text), we see that c_{n+1}'s interarrival time at stage three is

$$T_{n+1,3} = \begin{cases} T_{n+1,2} - W_n^{(2)} + S_{n+1,2}, & \text{if } T_{n+1,2} \geq W_n^{(2)} \\ S_{n+1,2}, & \text{if } T_{n+1,2} < W_n^{(2)}. \end{cases}$$



Similar to equation (4.3) in the text, c_{n+1}'s waiting time at the third stage is

$$W_{n+1}^{(3)} = \begin{cases} S_{n+1,3}, & \text{if } T_{n+1,3} \geq W_n^{(3)} \\ W_n^{(3)} - T_{n+1,3} + S_{n+1,3}, & \text{if } T_{n+1,3} < W_n^{(3)}. \end{cases}$$

Comparing $T_{n+1,2}$ and $T_{n+1,3}$ we have, in general, for c_{n+1}'s interarrival time at stage i, i=2,3,...,

$$T_{n+1,i} = \begin{cases} T_{n+1,i-1} - W_n^{(i-1)} + S_{n+1,i-1}, & \text{if } T_{n+1,i-1} \geq W_n^{(i-1)} \\ S_{n+1,i-1}, & \text{if } T_{n+1,i-1} < W_n^{(i-1)}. \end{cases}$$

Similarly, comparing $W_{n+1}^{(1)}$, $W_{n+1}^{(2)}$ and $W_{n+1}^{(3)}$ gives a general recursive formula for c_{n+1}'s waiting time at stage i, i=1,2,...,

$$W_{n+1}^{(i)} = \begin{cases} S_{n+1,i}, & \text{if } T_{n+1,i} \geq W_n^{(i)} \\ W_n^{(i)} - T_{n+1,i} + S_{n+1,i}, & \text{if } T_{n+1,i} < W_n^{(i)}. \end{cases}$$

Thus, we can obtain the recursive formulae for any number of queues in series.

TABLE 1. BIVARIATE DATA SETS

DESCRIPTION OF DATA	SAMPLE SIZE AND CORRELATION[1]	MARGINALS: MEAN AND VARIANCE	PARAMETERS[2] OF UNIVARIATE NEGATIVE BINOMIAL (EQ.(2.1)) (ν,θ)	FIT OF UNIVARIATE NEGATIVE BINOMIAL TO MARGINALS (χ^2,df,P[3])
ARBOUS-SICHEL ABSENTEEISM	248, 0.73	(X)1947 :4.70,18.66 (Y)1948 :4.48,18.66	1.58,2.99 1.56,2.88	8.9,13,0.78 11.0,12,0.53
BATES-NEYMAN DISEASES	1286, 0.42	(X)Dig.D. :1.40, 5.06 (Y)Res.D. :5.32,22.13	0.53,2.62 1.69,3.16	12.4,10,0.26 25.4,20,0.20
AIRCRAFT PARTS DEMAND	72, 0.54	(X)1st 3 mos:3.28,12.77 (Y)4th mo :0.65, 1.07	1.13,2.90 1.01,0.65	8.1, 6,0.23 1.0, 1,0.32
AIRCRAFT FLIGHT ABORTS	109,-0.16	(X)1st 6 mos:0.62, 1.03 (Y)2nd 6 mos:0.72, 1.08	0.95,0.66 1.51,0.48	0.24,1,0.62 6.3, 1,0.01
AIRCRAFT TOTAL ABORTS	203, 0.29	(X)A_1 :5.61,25.77 (Y)A_2 :6.00,23.85	1.56,3.60 2.01,2.98	20.1,15,0.17 11.9,15,0.69
AIRCRAFT TOTAL ABORTS	387, 0.23	(X)A_3 :7.07,45.04 (Y)A_4 :8.95,45.18	1.32,5.37 2.21,4.05	14.5,21,0.85 27.6,22,0.19
AIRCRAFT TOTAL ABORTS	428, 0.20	(X)A_4 :9.22,44.35 (Y)A_5 :8.78,48.31	2.42,3.81 2.15,4.09	19.9,25,0.75 23.2,25,0.56

[1]Pearson product-moment

[2]Via method of moments

[3]Probability of exceeding computed χ^2 value with indicated degrees of freedom

TABLE 2. G-B-N(1,θ,ν) MODEL[1]. OBSERVED AND EXPECTED CELL FREQUENCIES FOR ARBOUS-SICHEL DATA (248 WORKERS).

ABSENCES, 1948

ABSENCES, 1947	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27+
0	7 (11·90)	7 (6·04)	8 (4·43)	1 (2·25)	1 (1·11)	(0·53)	(0·25)	(0·11)	(0·05)	(0·02)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1	9 (8·04)	6 (6·66)	6 (6·70)	7 (4·43)	1 (2·63)	2 (1·45)	0·60	0·43	0·21	0·11	0·05	0·03	0·01	0·01	—	—	—	—	—	—	—	—	—	—	—	—	—	—
2	9 (4·33)	9 (6·76)	6 (6·03)	5 (5·20)	1 (3·60)	3 (2·30)	(1·46)	2 (0·85)	(0·48)	1 (0·25)	(0·14)	(0·07)	(0·4)	(0·02)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	—	—	—
3	2 (2·26)	8 (4·62)	9 (5·20)	2 (4·66)	4 (3·60)	3 (2·01)	(1·60)	(1·20)	3 (0·79)	1 (0·47)	(0·27)	(0·15)	(0·09)	(0·05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—	—
4	1 (1·11)	7 (2·63)	5 (3·60)	6 (3·60)	3 (3·04)	(2·60)	2 (2·34)	1 (1·80)	1 (1·08)	(0·80)	1 (0·42)	(0·25)	(0·15)	1 (0·09)	(0·05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—	—
5	2 (0·53)	1 (1·46)	1 (2·30)	2 (3·01)	5 (3·60)	2 (2·60)	1 (2·21)	1 (1·60)	(1·23)	1 (0·85)	3 (0·56)	(0·38)	(0·23)	(0·14)	(0·03)	1 (0·05)	(0·03)	(0·01)	(0·01)	(0·01)	—	—	—	—	—	—	—	—
6	(0·25)	1 (0·80)	1 (1·46)	2 (1·86)	2 (2·34)	3 (2·21)	1 (1·86)	2 (1·34)	(1·37)	2 (0·94)	(0·66)	1 (0·45)	(0·30)	1 (0·19)	(0·12)	(0·07)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	—	—	—	—	—	—	—
7	(0·11)	(0·02)	(0·05)	1 (1·20)	1 (1·60)	2 (1·60)	1 (1·44)	(1·46)	3 (1·21)	(0·95)	(0·71)	(0·51)	(0·36)	(0·24)	(0·16)	(0·10)	(0·06)	(0·04)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	—	—	—	—	—
8	(0·05)	(0·21)	1 (0·60)	(0·70)	2 (1·06)	2 (1·23)	(1·27)	(1·21)	2 (1·07)	(0·89)	1 (0·71)	1 (0·54)	(0·39)	(0·28)	(0·19)	(0·13)	(0·08)	(0·05)	(0·03)	1 (0·02)	(0·01)	(0·01)	(0·01)	—	—	—	—	—
9	(0·02)	(0·11)	(0·28)	(0·47)	1 (0·69)	(0·85)	3 (0·84)	(0·86)	2 (0·89)	(0·78)	(0·66)	(0·52)	2 (0·40)	(0·30)	(0·21)	(0·15)	(0·10)	(0·07)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—	—
10	(0·01)	(0·05)	(0·14)	1 (0·27)	(0·42)	(0·66)	1 (0·66)	(0·71)	1 (0·71)	(0·66)	(0·57)	1 (0·48)	2 (0·39)	(0·30)	(0·22)	1 (0·16)	(0·12)	(0·08)	(0·05)	(0·04)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—
11	(0·01)	(0·02)	(0·07)	(0·15)	1 (0·25)	(0·36)	(0·45)	(0·51)	(0·54)	(0·52)	(0·48)	(0·43)	1 (0·35)	(0·29)	(0·22)	(0·17)	(0·13)	(0·09)	1 (0·06)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—
12	—	(0·01)	(0·04)	(0·08)	(0·16)	1 (0·23)	(0·30)	(0·36)	1 (0·39)	1 (0·40)	1 (0·39)	(0·35)	(0·32)	(0·26)	1 (0·21)	(0·16)	(0·12)	(0·09)	(0·07)	(0·05)	(0·03)	(0·02)	(0·01)	(0·01)	(0·01)	—	—	—
13	—	(0·01)	(0·03)	(0·05)	(0·09)	1 (0·14)	(0·19)	(0·24)	(0·28)	1 (0·30)	1 (0·30)	1 (0·29)	1 (0·26)	1 (0·23)	(0·19)	(0·15)	(0·12)	(0·09)	(0·07)	(0·05)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—	—
14	—	—	(0·01)	(0·03)	1 (0·05)	(0·08)	(0·12)	1 (0·16)	(0·19)	(0·21)	(0·22)	(0·22)	(0·21)	1 (0·19)	(0·19)	(0·14)	(0·11)	(0·09)	(0·07)	1 (0·05)	(0·04)	(0·03)	(0·02)	(0·01)	(0·01)	—	—	—
15	—	—	(0·01)	(0·01)	(0·03)	(0·05)	(0·07)	1 (0·10)	(0·13)	(0·15)	(0·16)	(0·17)	(0·16)	(0·15)	1 (0·14)	(0·12)	(0·10)	(0·08)	(0·07)	(0·05)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—
16	—	—	—	(0·01)	(0·01)	(0·03)	(0·04)	(0·06)	(0·08)	(0·10)	1 (0·12)	(0·13)	(0·12)	(0·12)	(0·11)	(0·10)	1 (0·10)	(0·08)	(0·08)	(0·05)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	—	—
17	—	—	—	(0·01)	(0·01)	(0·01)	(0·03)	(0·04)	(0·05)	(0·07)	1 (0·08)	(0·09)	(0·09)	(0·09)	1 (0·09)	(0·08)	(0·08)	(0·08)	(0·05)	(0·04)	(0·04)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)
18	—	—	—	—	(0·01)	(0·01)	(0·02)	(0·03)	(0·03)	(0·04)	(0·05)	(0·06)	(0·07)	(0·07)	(0·07)	(0·07)	(0·06)	(0·05)	(0·05)	(0·04)	(0·03)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)
19	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·02)	(0·03)	(0·04)	(0·04)	(0·05)	(0·05)	(0·05)	(0·05)	(0·05)	(0·04)	(0·04)	(0·04)	(0·03)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)
20	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·03)	(0·03)	(0·03)	(0·04)	(0·04)	(0·04)	(0·04)	(0·03)	(0·03)	(0·03)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)
21	—	—	—	—	—	—	—	1 (0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)	1 (0·03)	(0·03)	(0·03)	(0·03)	(0·02)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)
22	—	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	1 (0·01)	(0·01)
23	—	—	—	—	—	—	—	—	—	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·02)	(0·02)	(0·02)	(0·02)	(0·02)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)	(0·01)

[1]Guldberg-Bates-Neyman bivariate negative binomial distribution of (2.6). ML estimates are: $\hat{\theta}=2.8956$, $\hat{\nu}=1.5854$. There results $\chi^2=17.0$ and $P\doteq 0.20$ for df=13.

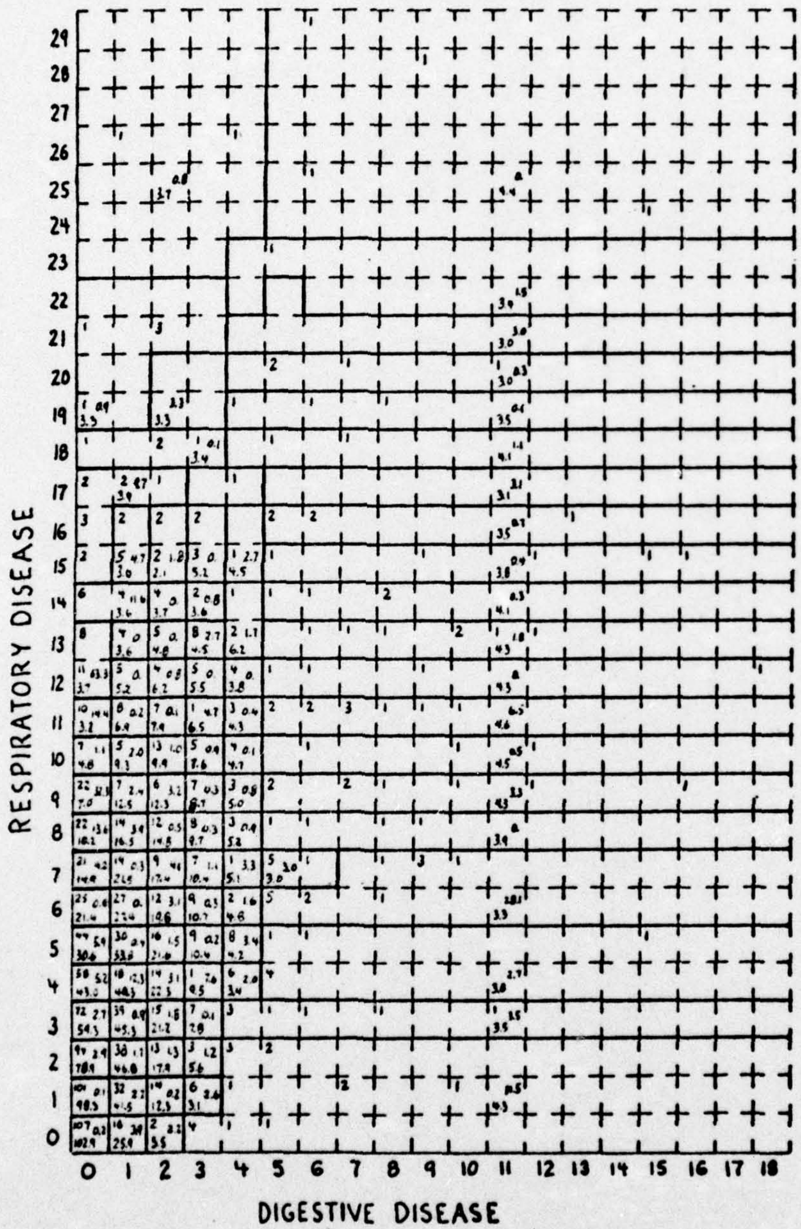
TABLE 3. BNB(a, 0, 0, p, q, ν) MODEL[1]. EXPECTED CELL FREQUENCIES FOR ARBOUS-SICHEL DATA (248 WORKERS)

ABSENCES, 1948 (columns); ABSENCES, 1947 (rows)

	0					5					10				
0	11.51	7.92	4.49	2.36	1.19	0.59	0.29	0.14	0.06	0.03	0.01	0.01	0.00	0.00	0.00
	8.25	8.54	6.46	4.25	2.58	1.48	0.82	0.44	0.23	0.12	0.06	0.03	0.02	0.01	0.00
	4.86	6.73	6.31	4.94	3.47	2.27	1.41	0.84	0.49	0.27	0.15	0.08	0.04	0.02	0.01
	2.66	4.61	5.14	4.65	3.71	2.71	1.85	1.21	0.76	0.46	0.27	0.16	0.09	0.05	0.03
	1.40	2.91	3.76	3.86	3.43	2.76	2.06	1.46	0.99	0.64	0.41	0.25	0.15	0.09	0.05
5	0.72	1.74	2.56	2.93	2.87	2.52	2.04	1.55	1.12	0.78	0.52	0.34	0.22	0.13	0.08
	0.36	1.01	1.66	2.09	2.24	2.13	1.85	1.51	1.16	0.86	0.61	0.42	0.28	0.18	0.11
	0.18	0.57	1.03	1.42	1.65	1.69	1.57	1.36	1.11	0.87	0.65	0.47	0.33	0.22	0.15
	0.09	0.31	0.62	0.93	1.16	1.27	1.26	1.16	1.00	0.82	0.64	0.48	0.35	0.25	0.17
	0.04	0.17	0.36	0.59	0.79	0.92	0.97	0.94	0.85	0.73	0.60	0.47	0.36	0.27	0.19
10	0.02	0.09	0.21	0.36	0.52	0.64	0.71	0.73	0.70	0.63	0.54	0.44	0.35	0.27	0.20
	0.01	0.05	0.12	0.22	0.33	0.43	0.51	0.55	0.55	0.51	0.46	0.39	0.32	0.26	0.20
	0.01	0.02	0.07	0.13	0.21	0.29	0.35	0.40	0.42	0.41	0.38	0.34	0.29	0.24	0.19
	0.00	0.01	0.04	0.08	0.13	0.19	0.24	0.28	0.31	0.31	0.30	0.28	0.25	0.21	0.17
	0.00	0.01	0.02	0.04	0.08	0.12	0.16	0.20	0.22	0.24	0.24	0.23	0.21	0.18	0.16
15	0.00	0.00	0.01	0.02	0.05	0.07	0.10	0.13	0.16	0.17	0.18	0.18	0.17	0.15	0.13
	0.00	0.00	0.01	0.01	0.03	0.04	0.07	0.09	0.11	0.12	0.13	0.13	0.13	0.12	0.11
	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.06	0.07	0.09	0.10	0.10	0.10	0.10	0.09
	0.00	0.00	0.00	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.07	0.08	0.08	0.07
	0.00	0.00	0.00	0.00	0.01	0.01	0.02	0.02	0.03	0.04	0.05	0.05	0.06	0.06	0.06
20	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.03	0.03	0.04	0.04	0.04	0.04
	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.02	0.02	0.03	0.03	0.03	0.03

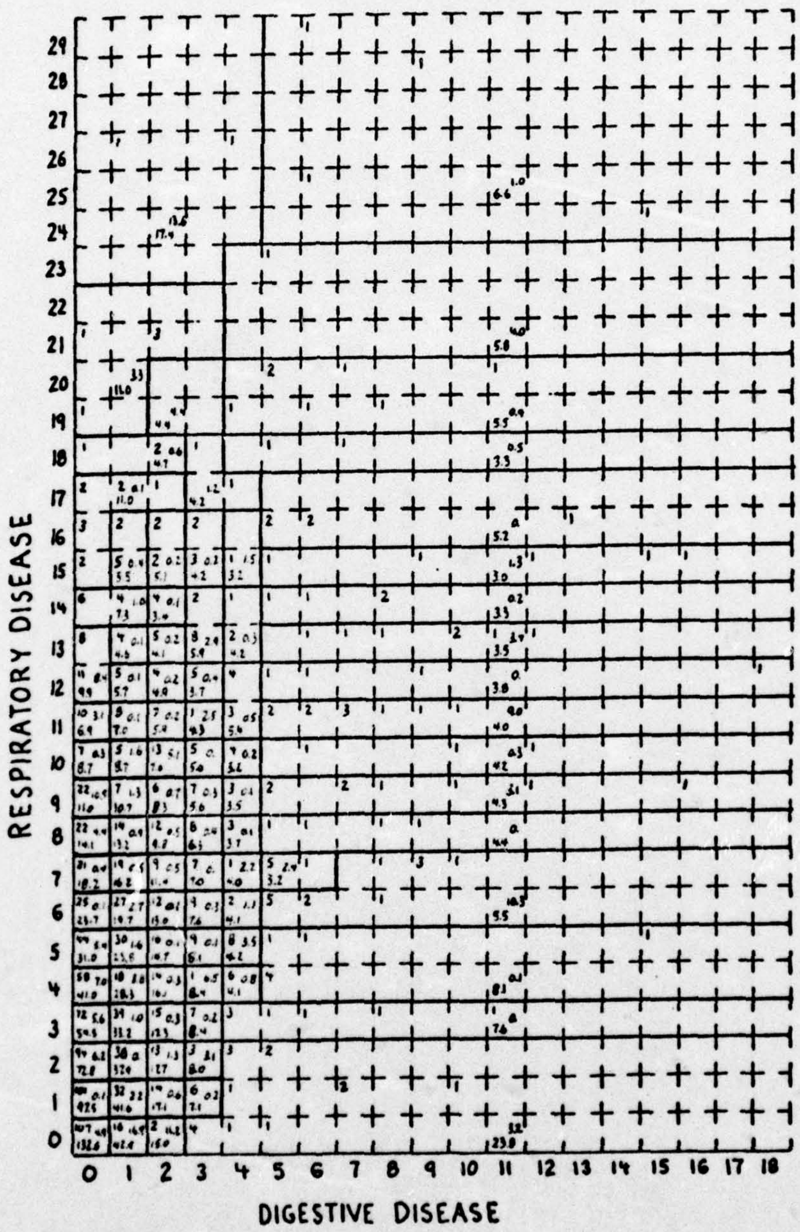
[1] Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}=0.0370$, $\hat{p}=0.1010$, $\hat{q}=0.0970$, $\hat{\nu}=1.5480$.

TABLE 4. G-B-N(α, θ, ν) MODEL[1].
OBSERVED AND EXPECTED CELL FREQUENCIES
FOR BATES-NEYMAN DATA (1286 WORKERS).



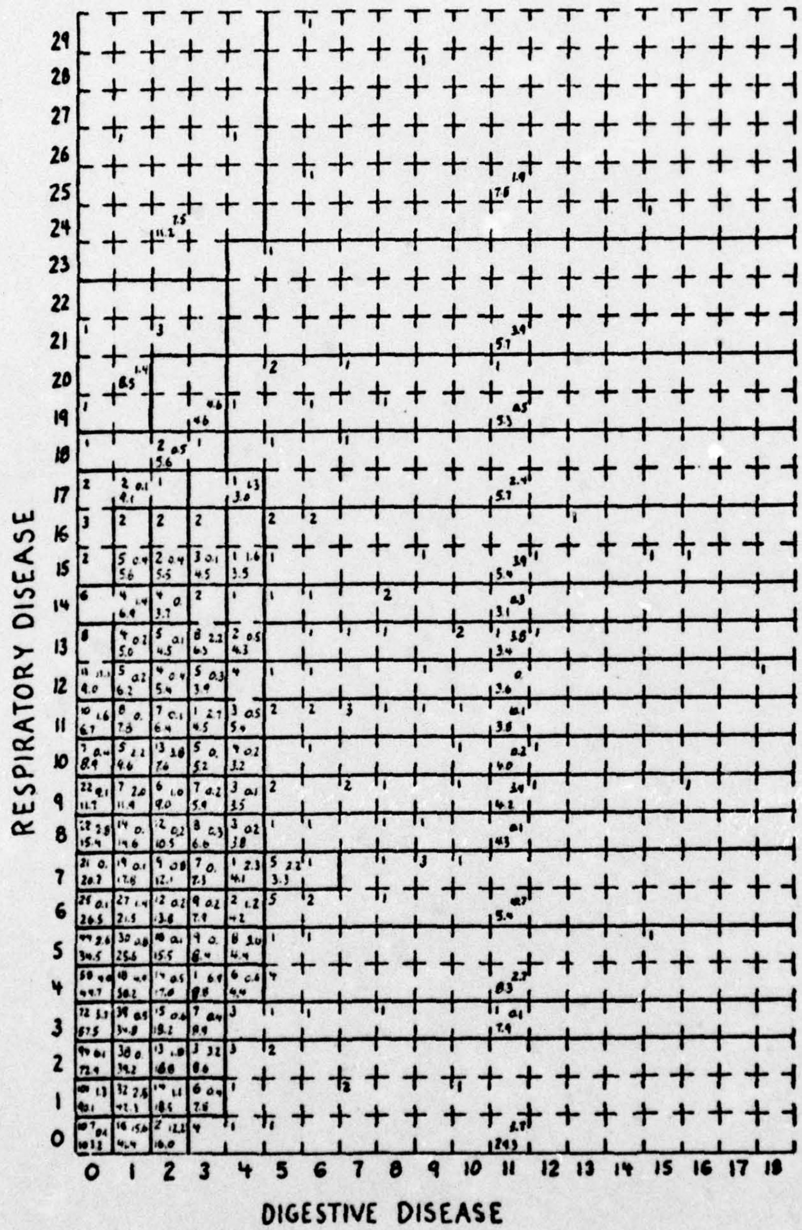
[1] Guldberg-Bates-Neyman bivariate negative binomial distribution of (2.6). ML estimates are: $\hat{\alpha}=3.798$, $\hat{\theta}=0.952$, $\hat{\nu}=1.471$. There results $\chi^2=353.8$ and $P \doteq 0$ for df=94.

TABLE 5. BNB(a,b,c,p,q,1) MODEL[1]. OBSERVED AND EXPECTED CELL FREQUENCIES FOR BATES-NEYMAN DATA (1286 WORKERS).



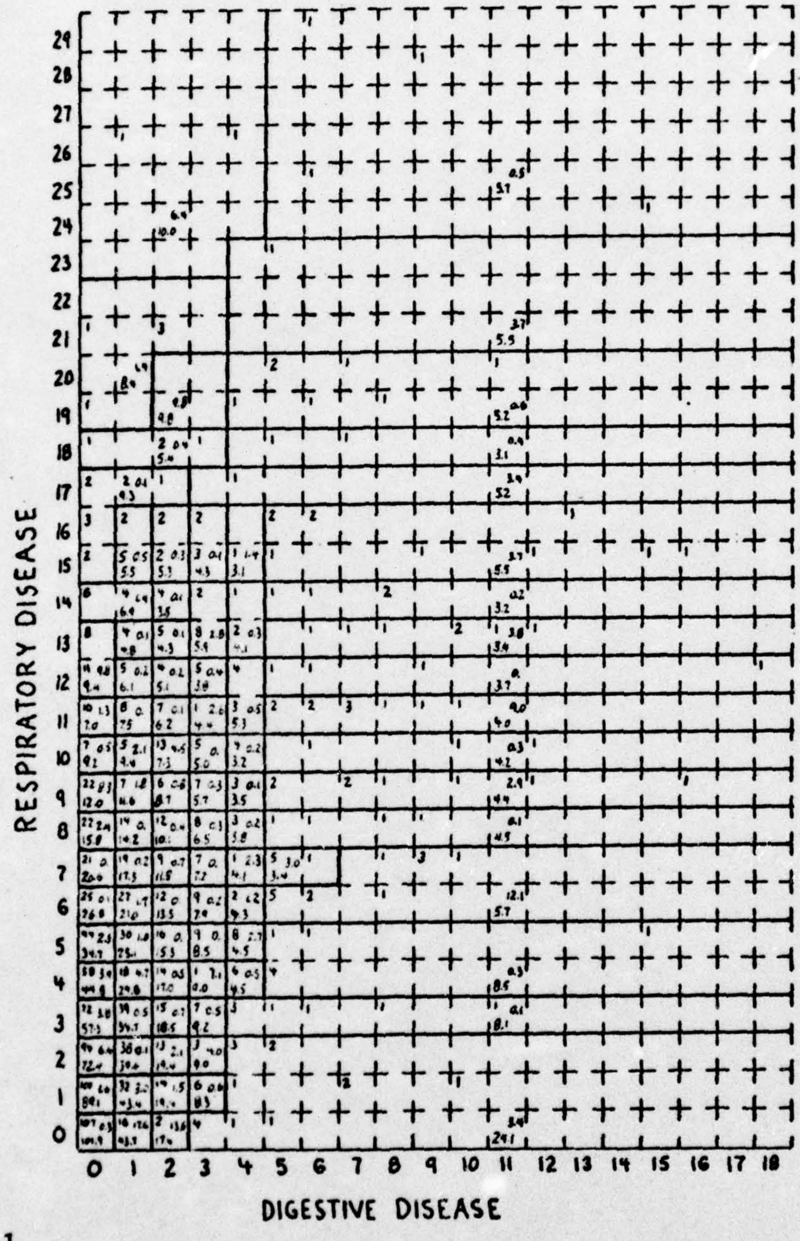
[1]Paulson-Uppuluri bivariate geometric distribution of (2.36). ML estimates are: $\hat{a}=0.1338$, $\hat{b}=0.0191$, $\hat{c}=0.0330$, $\hat{p}=0.1860$, $\hat{q}=0.4516$. There results $\chi^2=202.4$ and $P\doteq 0$ for df=87.

TABLE 6. BNB(a,b,0,p,q,ν) MODEL[1].
OBSERVED AND EXPECTED CELL FREQUENCIES
FOR BATES-NEYMAN DATA (1286 WORKERS).



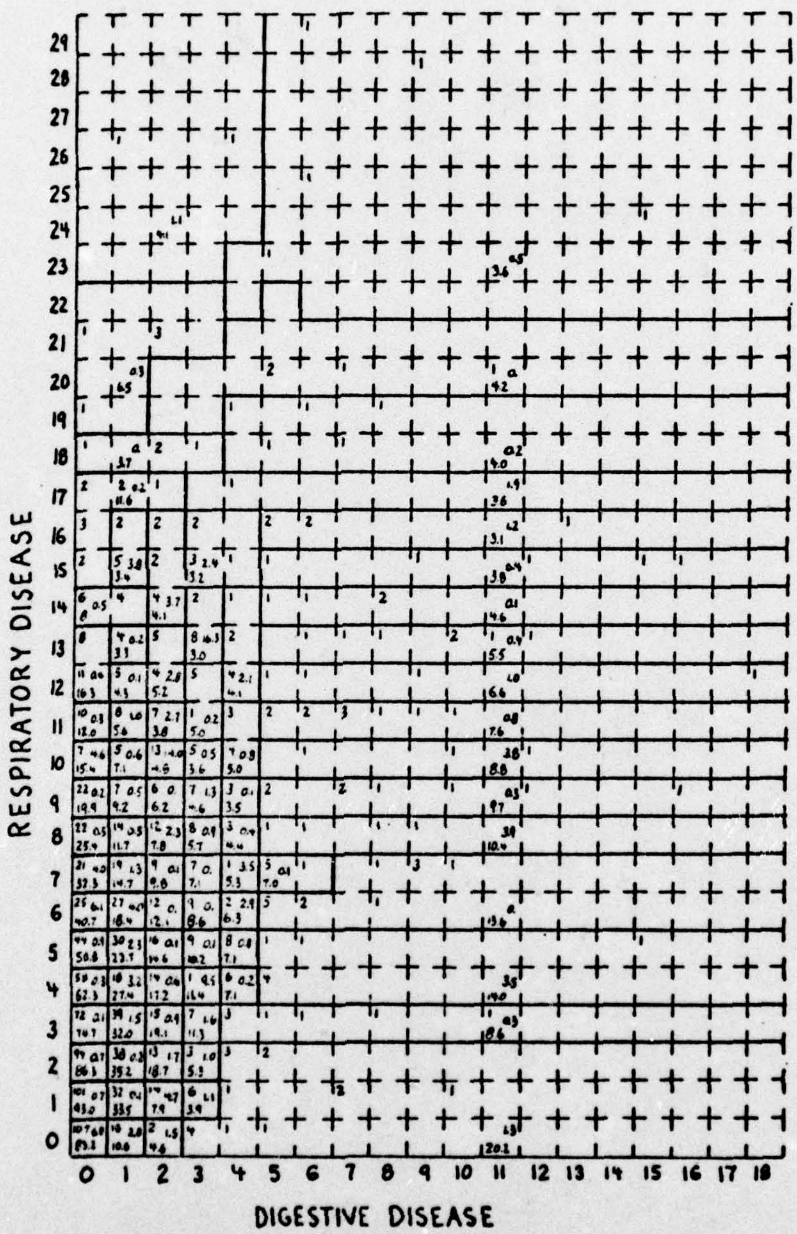
[1]New bivariate negative binomial distribution of (2.37). ML estimates are: $\hat{a}$=0.2844, $\hat{b}$=0.0256, $\hat{p}$=0.2537, $\hat{q}$=0.5880, $\hat{\nu}$=1.1687. There results χ^2= 194.9 and $P \doteq 0$ for df=86.

TABLE 7. BNB(α,0,0,p,q,ν) MODEL[1]. OBSERVED AND EXPECTED CELL FREQUENCIES FOR BATES-NEYMAN DATA (1286 WORKERS).



[1]Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}$=0.3725, $\hat{p}$=0.3086, $\hat{q}$=0.6299, $\hat{\nu}$=1.1654. There results χ^2=195.4 and P$\doteq$0 for df=89.

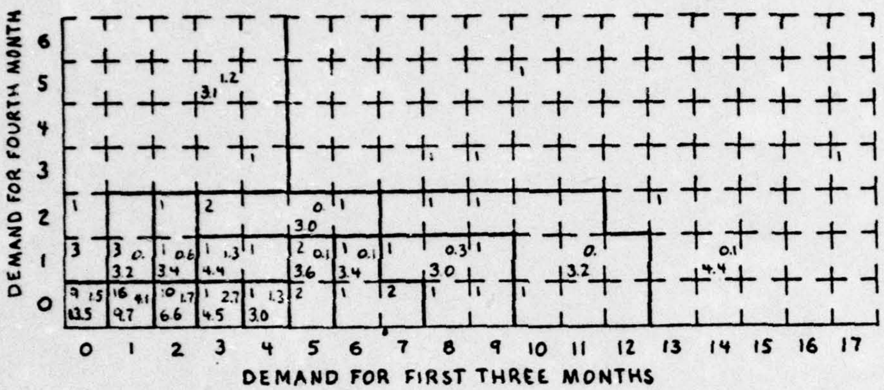
TABLE 8. BNB-TR(θ,ν_1,ν_2,ν_3) MODEL[1].
OBSERVED AND EXPECTED CELL FREQUENCIES
FOR BATES-NEYMAN DATA (1286 WORKERS).



[1]Bivariate negative binomial distribution via a tri-variate reduction - (2.57). ML estimates are: $\hat{\theta}=2.8375, \hat{\nu}_1=0.1717, \hat{\nu}_2=0.3529, \hat{\nu}_3=1.5115$. There results $\chi^2=152.6$ and $P \doteq 0$ for df=82.

TABLE 9. OBSERVED AND EXPECTED CELL FREQUENCIES OF DEMAND FOR 72 AIRCRAFT PARTS.

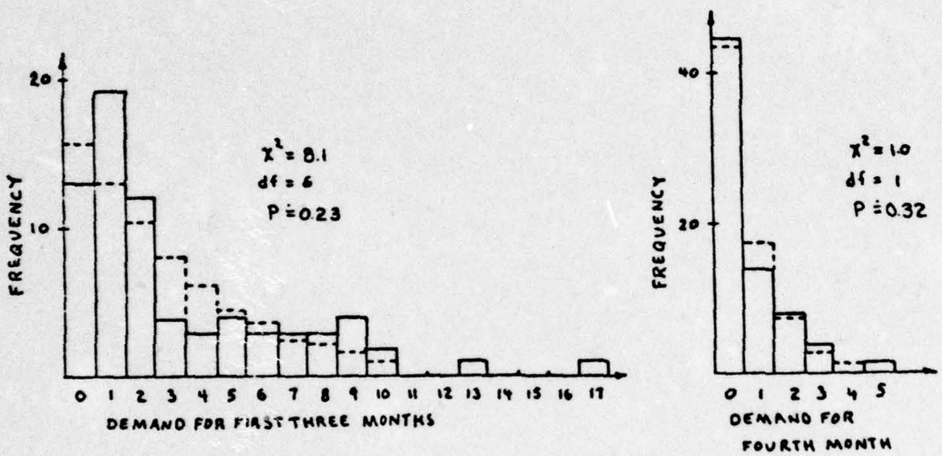
(a) BIVARIATE, G-B-N$(\alpha,\theta,\nu)^1$



$\chi^2 = 15.0$
df = 11
$P \doteq 0.18$

1-GOLDBERG-BATES-NEYMAN OF (2.6). MLE: $\hat{\alpha} = 0.199$, $\hat{\theta} = 2.987$, $\hat{\nu} = 1.100$

(b) UNIVARIATES

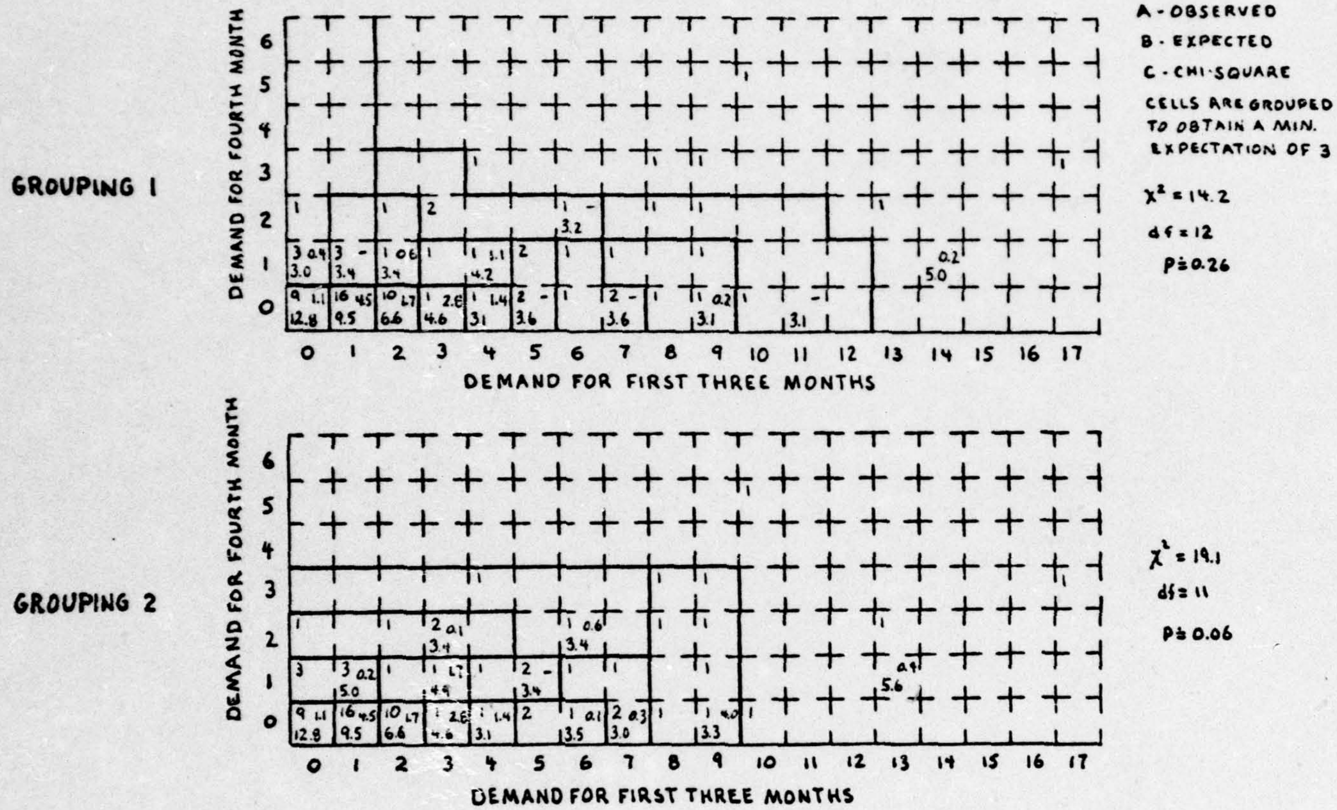


$\chi^2 = 8.1$
df = 6
$P \doteq 0.23$

$\chi^2 = 1.0$
df = 1
$P \doteq 0.32$

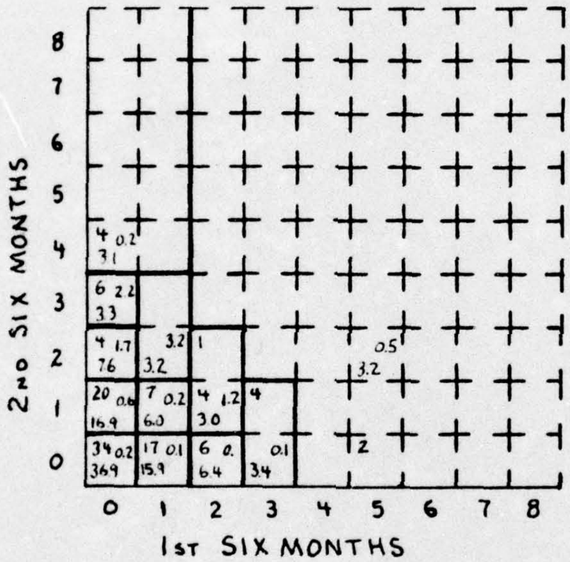
— OBSERVED
--- EXPECTED (NEGATIVE BINOMIAL)

TABLE 10. TWO GROUPINGS OF CELLS.
OBSERVED AND EXPECTED[1] CELL FREQUENCIES
OF DEMAND FOR 72 AIRCRAFT PARTS.



[1] Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}=0.1024$, $\hat{p}=0.2307$, $\hat{q}=0.0564$, $\hat{\nu}=1.1200$.

TABLE II. BNB(a,b,c,p,q,1) MODEL[1]. OBSERVED AND EXPECTED CELL FREQUENCIES OF FLIGHT ABORTS FOR 109 AIRCRAFT.



[1]Paulson-Uppuluri bivariate geometric distribution of (2.36). ML estimates are: $\hat{a}=0$, $\hat{b}=0.6820$, $\hat{c}=0.3179$, $\hat{p}=0.1655$ $\hat{q}=0.3299$. There results $\chi^2=10.2$ and $P \doteq 0.12$ for df=6.

TABLE 12. OBSERVED CELL FREQUENCIES OF TOTAL ABORTS FOR 203 AIRCRAFT. NO INTERVENING OVERHAUL.

TOTAL ABORTS FOR PERIOD 2, A_2 \ TOTAL ABORTS FOR PERIOD 1, A_1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	1	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	1	0	0	1	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	3	0	2	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	4	3	1	0	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	2	1	1	2	1	3	2	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	3	2	1	1	1	0	1	1	0	2	3	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5	1	0	2	2	1	3	0	2	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	3	3	0	1	2	3	0	0	1	1	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
	3	4	4	1	3	3	1	0	4	2	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	3	4	3	0	0	2	0	4	2	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	3	2	2	2	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	1	1	1	3	0	1	1	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

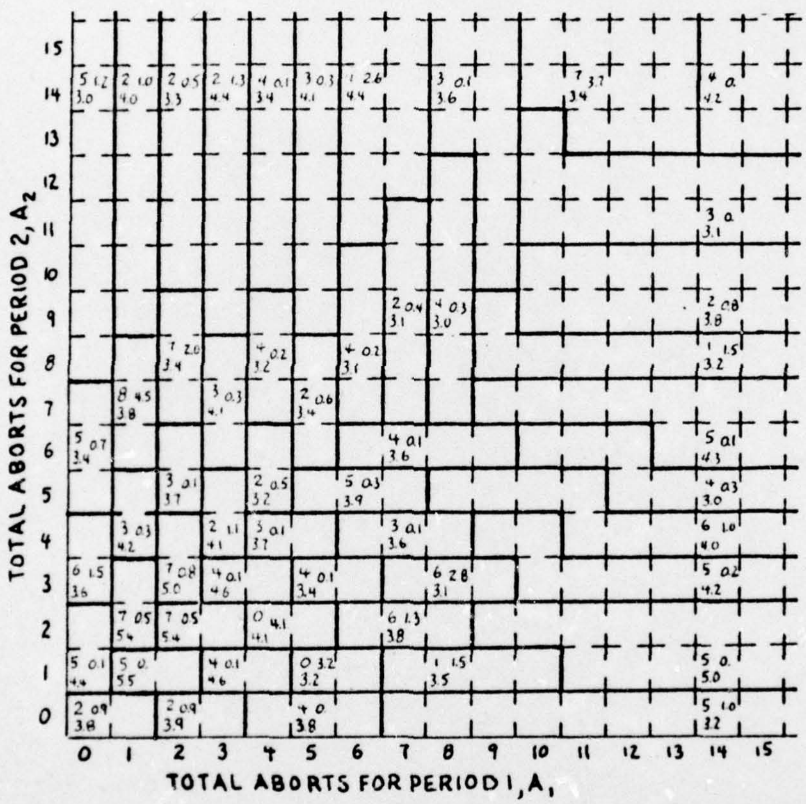
TABLE 13. BNB(a,0,0,p,q,ν) MODEL[1]. EXPECTED CELL FREQUENCIES OF TOTAL ABORTS FOR 203 AIRCRAFT.

TOTAL ABORTS FOR PERIOD 2, A_2															
	0.02	0.03	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.04	0.04	0.03	0.03	0.03
20	0.03	0.04	0.06	0.06	0.07	0.07	0.07	0.07	0.06	0.06	0.05	0.05	0.04	0.04	0.03
	0.04	0.06	0.07	0.08	0.09	0.09	0.09	0.08	0.08	0.07	0.06	0.06	0.05	0.04	0.04
	0.05	0.08	0.10	0.11	0.11	0.11	0.11	0.10	0.09	0.09	0.08	0.07	0.06	0.05	0.04
	0.06	0.10	0.13	0.14	0.14	0.14	0.14	0.13	0.12	0.11	0.09	0.08	0.07	0.05	0.05
	0.08	0.13	0.16	0.18	0.18	0.19	0.17	0.16	0.14	0.13	0.11	0.10	0.09	0.07	0.06
15	0.11	0.17	0.21	0.23	0.23	0.23	0.21	0.20	0.18	0.16	0.14	0.12	0.10	0.09	0.07
	0.15	0.23	0.27	0.29	0.29	0.28	0.26	0.24	0.22	0.19	0.17	0.14	0.12	0.10	0.09
	0.19	0.30	0.35	0.37	0.37	0.35	0.33	0.29	0.26	0.23	0.20	0.17	0.14	0.12	0.10
	0.25	0.38	0.45	0.47	0.46	0.43	0.40	0.36	0.32	0.27	0.24	0.20	0.17	0.14	0.12
	0.33	0.49	0.57	0.59	0.57	0.53	0.49	0.43	0.38	0.33	0.28	0.23	0.20	0.16	0.13
10	0.43	0.63	0.72	0.73	0.71	0.65	0.59	0.52	0.45	0.38	0.32	0.27	0.23	0.19	0.15
	0.56	0.80	0.90	0.91	0.86	0.79	0.71	0.62	0.53	0.45	0.38	0.31	0.26	0.21	0.17
	0.71	1.01	1.12	1.12	1.05	0.95	0.84	0.73	0.62	0.52	0.43	0.36	0.29	0.24	0.19
	0.91	1.27	1.38	1.36	1.26	1.13	0.98	0.84	0.71	0.59	0.49	0.40	0.32	0.25	0.21
	1.14	1.56	1.68	1.63	1.49	1.32	1.14	0.96	0.80	0.66	0.54	0.44	0.35	0.28	0.22
5	1.40	1.90	2.00	1.92	1.74	1.52	1.29	1.08	0.89	0.73	0.59	0.47	0.38	0.30	0.23
	1.69	2.25	2.34	2.21	1.97	1.70	1.43	1.18	0.97	0.78	0.62	0.49	0.39	0.30	0.24
	1.96	2.58	2.64	2.45	2.16	1.84	1.53	1.25	1.01	0.80	0.63	0.50	0.39	0.30	0.23
	2.17	2.80	2.83	2.59	2.24	1.88	1.54	1.24	0.99	0.78	0.61	0.47	0.36	0.28	0.21
	2.18	2.77	2.75	2.47	2.11	1.74	1.41	1.12	0.88	0.68	0.52	0.40	0.31	0.23	0.17
0	1.71	2.14	2.08	1.85	1.55	1.26	1.00	0.78	0.60	0.46	0.35	0.26	0.20	0.15	0.11
	0					5					10				

TOTAL ABORTS FOR PERIOD 1, A_1

[1]Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}=0.6795$, $\hat{p}=0.6827$, $\hat{q}=0.6971$, $\hat{\nu}=1.7701$.

TABLE 14. BNB(a,0,0,p,q,ν) MODEL[1].
OBSERVED AND EXPECTED CELL FREQUENCIES
OF TOTAL ABORTS FOR 203 AIRCRAFT.
NO INTERVENING OVERHAUL.



[1]Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}$=0.6795, $\hat{p}$=0.6827, $\hat{q}$=0.6971, $\hat{\nu}$=1.7701. There results χ^2=46.1 and P$\doteq$ 0.65 for df=50.

TABLE 15. OBSERVED CELL FREQUENCIES OF TOTAL ABORTS FOR 387 AIRCRAFT. INTERVENING OVERHAUL.

TOTAL ABORTS FOR PERIOD 4, A_4	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	1	0	0	1	1	0	0	2	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	2	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	1	0	1	2	0	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	2	1	0	2	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0
15	0	0	1	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	1	1	1	0	1	0	2	0	1	0	0	0	1	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	2	4	1	1	0	1	2	2	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	3	3	4	1	1	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	0	4	1	1	1	0	3	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	3	4	2	1	0	2	0	2	1	2	1	2	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	3	1	2	1	2	0	0	2	0	2	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	3	0	2	3	0	1	0	2	1	0	0	1	0	0	1	1	0	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	2	5	4	3	3	0	1	0	0	0	1	3	2	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	5	1	2	4	3	6	1	3	1	4	1	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
5	1	1	2	4	1	2	4	2	2	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	2	4	2	2	0	1	3	1	3	2	1	0	0	3	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1
	3	2	4	1	3	2	1	2	0	1	0	0	2	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	5	6	3	2	2	4	3	0	3	0	1	2	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	0	2	1	1	2	2	3	1	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	3	1	1	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TOTAL ABORTS FOR PERIOD 3, A_3

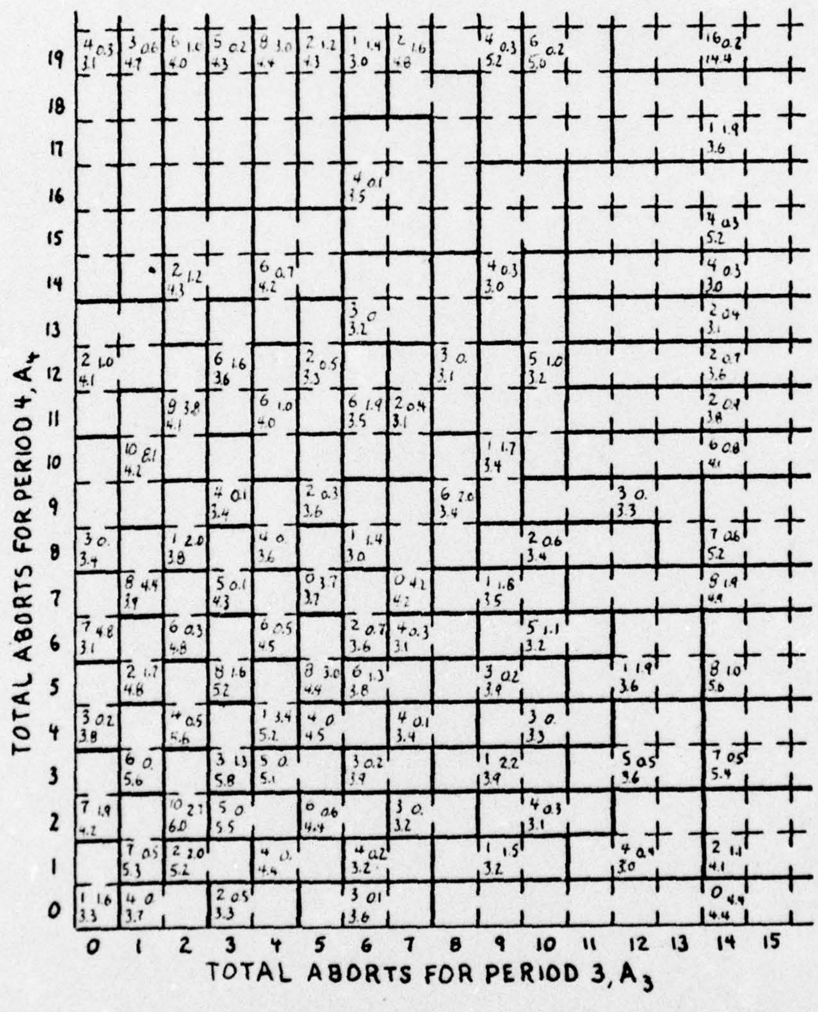
TABLE 16. BNB(a,o,o,p,q,ν) MODEL[1]. EXPECTED CELL FREQUENCIES OF TOTAL ABORTS FOR 387 AIRCRAFT.

TOTAL ABORTS FOR PERIOD 4, A_4															
	0.15	0.23	0.27	0.29	0.30	0.30	0.29	0.27	0.25	0.23	0.21	0.19	0.17	0.15	0.13
20	0.18	0.27	0.32	0.35	0.35	0.35	0.33	0.31	0.29	0.26	0.24	0.21	0.19	0.17	0.15
	0.21	0.32	0.38	0.41	0.41	0.41	0.39	0.36	0.33	0.30	0.27	0.24	0.22	0.19	0.17
	0.26	0.38	0.45	0.48	0.49	0.47	0.45	0.42	0.38	0.35	0.31	0.28	0.24	0.21	0.19
	0.31	0.46	0.53	0.56	0.57	0.55	0.52	0.48	0.44	0.40	0.35	0.31	0.28	0.24	0.21
	0.36	0.54	0.63	0.66	0.66	0.63	0.60	0.55	0.50	0.45	0.40	0.35	0.31	0.27	0.23
15	0.43	0.64	0.74	0.77	0.76	0.73	0.69	0.63	0.57	0.51	0.45	0.40	0.35	0.30	0.26
	0.51	0.75	0.86	0.90	0.88	0.84	0.78	0.72	0.65	0.58	0.51	0.45	0.39	0.33	0.29
	0.61	0.88	1.00	1.04	1.02	0.97	0.89	0.81	0.73	0.65	0.57	0.50	0.43	0.37	0.32
	0.71	1.03	1.17	1.20	1.17	1.10	1.02	0.92	0.82	0.73	0.63	0.55	0.48	0.41	0.35
	0.84	1.20	1.35	1.38	1.34	1.25	1.15	1.03	0.92	0.81	0.70	0.61	0.52	0.45	0.38
10	0.97	1.39	1.55	1.57	1.52	1.42	1.29	1.16	1.02	0.89	0.77	0.67	0.57	0.49	0.41
	1.13	1.59	1.77	1.79	1.71	1.59	1.44	1.28	1.13	0.98	0.85	0.73	0.62	0.52	0.44
	1.30	1.82	2.01	2.01	1.92	1.77	1.59	1.41	1.24	1.07	0.92	0.78	0.66	0.56	0.47
	1.48	2.06	2.25	2.25	2.13	1.95	1.75	1.54	1.34	1.15	0.98	0.84	0.70	0.59	0.49
	1.66	2.30	2.50	2.48	2.33	2.12	1.89	1.65	1.43	1.23	1.04	0.88	0.74	0.62	0.51
5	1.84	2.53	2.73	2.69	2.51	2.27	2.01	1.75	1.51	1.28	1.08	0.91	0.76	0.63	0.52
	1.99	2.72	2.92	2.85	2.65	2.38	2.09	1.81	1.55	1.31	1.10	0.92	0.76	0.63	0.52
	2.09	2.83	3.02	2.93	2.70	2.41	2.11	1.81	1.54	1.30	1.08	0.90	0.74	0.61	0.50
	2.09	2.81	2.97	2.86	2.62	2.33	2.02	1.73	1.46	1.22	1.01	0.83	0.68	0.56	0.45
	1.90	2.53	2.66	2.55	2.32	2.04	1.76	1.49	1.25	1.04	0.86	0.70	0.57	0.46	0.37
0	1.36	1.80	1.88	1.78	1.61	1.41	1.21	1.02	0.85	0.70	0.57	0.46	0.38	0.30	0.24
	0					5					10				

TOTAL ABORTS FOR PERIOD 3, A_3

[1]Clark's bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}=0.7470$, $\hat{p}=0.7526$, $\hat{q}=0.7939$, $\hat{\nu}=1.7359$.

TABLE 17. BNB(a,0,0,p,q,ν) MODEL[1].
OBSERVED AND EXPECTED CELL FREQUENCIES
OF TOTAL ABORTS FOR 387 AIRCRAFT.
INTERVENING OVERHAUL.

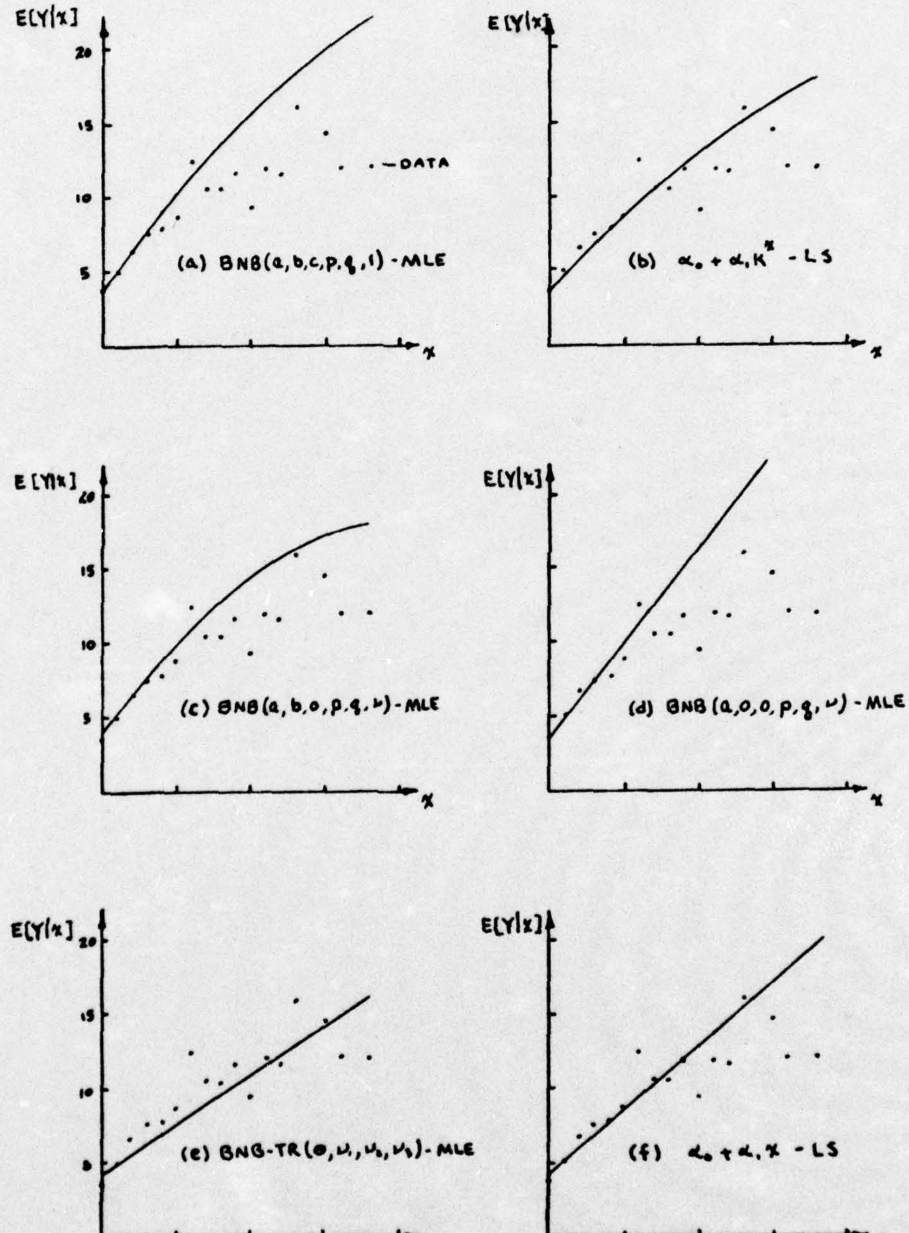


[1]Clark's, bivariate negative binomial distribution of (2.26). ML estimates are: $\hat{a}$=0.7470, $\hat{p}$=0.7526, $\hat{q}$=0.7939, $\hat{\nu}$=1.7359. There results χ^2=107.0 and P≐0.12 for df=91.

TABLE 18. OBSERVED CELL FREQUENCIES OF TOTAL ABORTS FOR 428 AIRCRAFT. FIRST TWO PERIODS AFTER OVERHAUL.

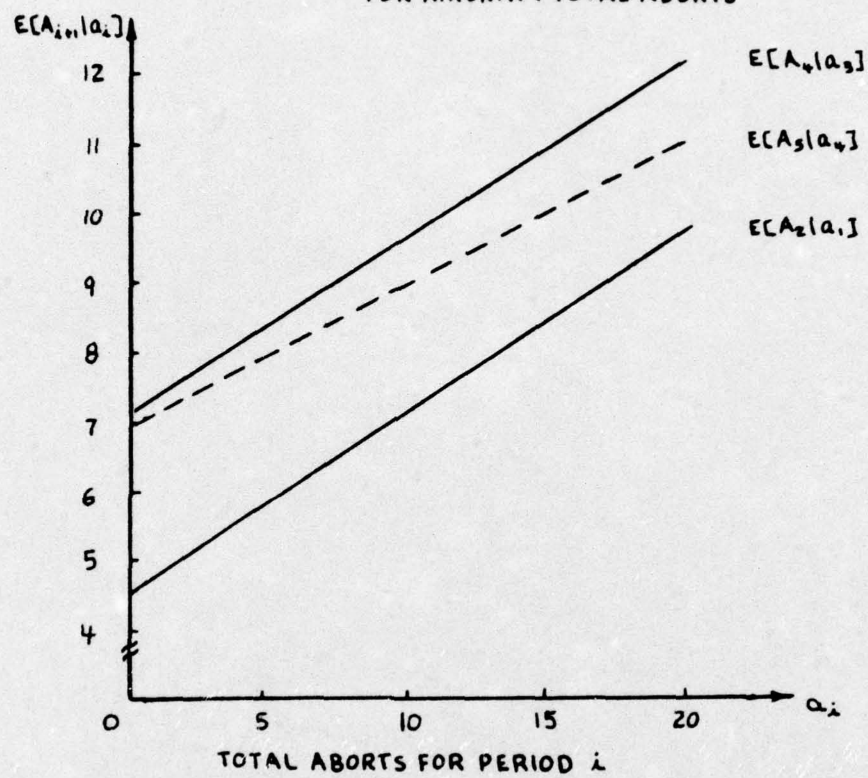
TOTAL ABORTS FOR PERIOD 5, A_5 \ TOTAL ABORTS FOR PERIOD 4, A_4	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34
	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	0	0	0	0	2	1	1	0	1	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	2	0	0	1	0	0	0	0	1	1	0	0	0	0	2	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	3	1	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	1	3	0	0	1	0	0	0	2	0	0	0	1	0	0	0	0	0	3	0	0	0	0	0	0	0	0	1	0	0	0	0	0
15	0	1	1	0	1	0	1	2	0	2	0	0	1	2	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	1	0	1	1	1	1	1	2	0	1	2	0	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	1	0	1	1	0	2	0	2	1	2	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
	1	1	0	0	3	0	0	0	2	1	1	1	0	0	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	2	0	2	1	3	1	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0
10	0	2	1	0	2	0	0	1	1	1	0	0	2	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	3	1	1	5	1	4	0	1	0	1	1	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	3	1	0	0	4	3	1	1	1	1	0	0	0	0	2	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	3	4	3	1	3	3	1	0	1	0	0	2	1	2	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	0	0
	0	4	3	2	1	2	1	2	1	1	0	0	0	2	1	0	4	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0
5	1	1	2	2	3	1	5	1	3	5	1	2	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	2	3	0	1	1	4	3	2	4	2	3	4	1	1	2	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	1	2	5	0	6	3	1	1	3	2	3	1	0	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	2	6	2	2	2	1	2	1	0	1	1	0	0	1	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	1	3	1	2	4	2	2	0	2	2	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
0	0	1	2	2	1	2	1	2	1	2	1	1	0	0	2	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURE 1. REGRESSION FUNCTIONS, E[Y|x], FOR BATES-NEYMAN DATA



X - DIGESTIVE DISEASE
Y - RESPIRATORY DISEASE

FIGURE 2. REGRESSION FUNCTIONS, $E[A_{i+1} | a_i]$, FOR AIRCRAFT TOTAL ABORTS



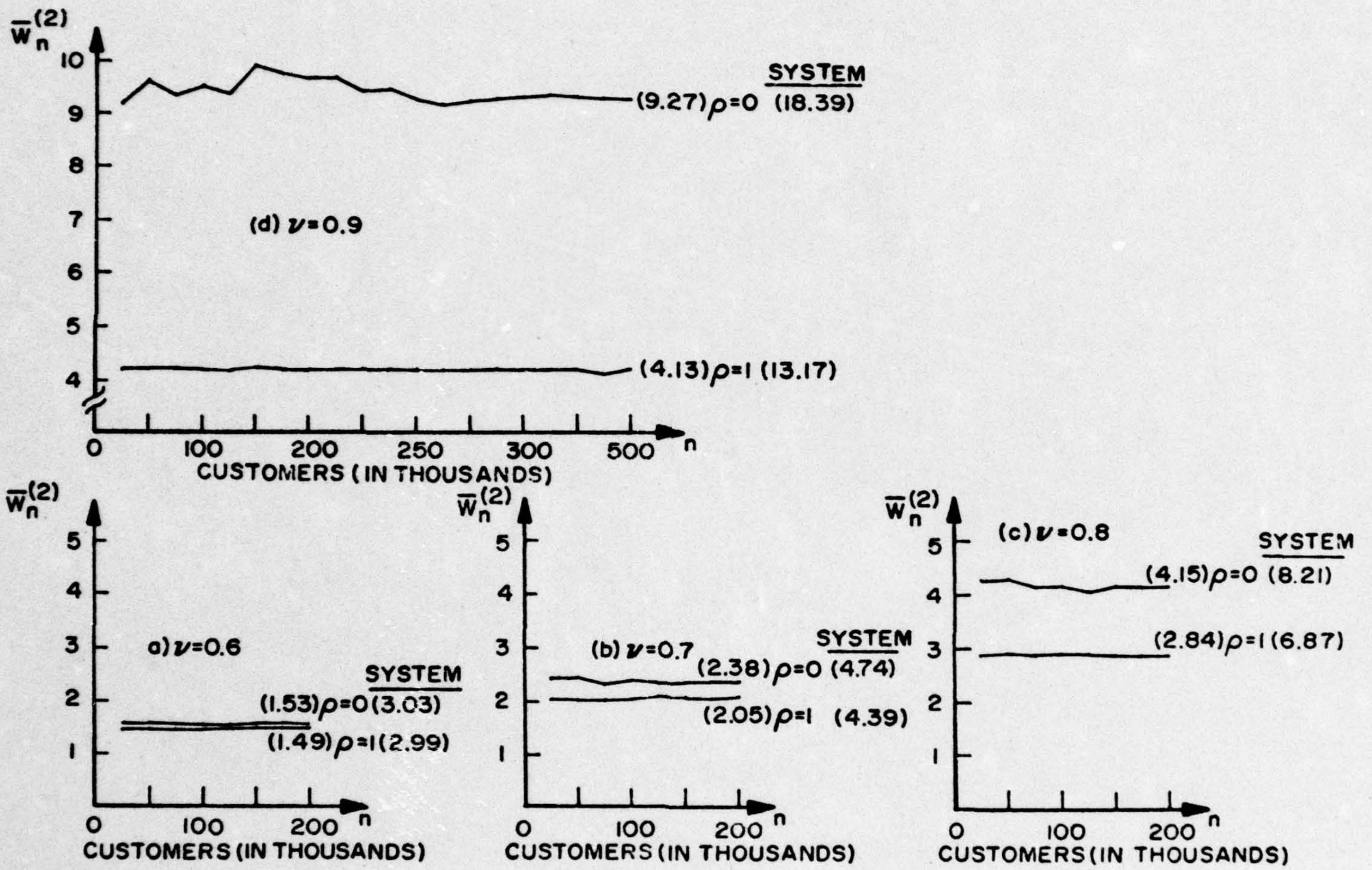


FIGURE 3. MEAN WAITING TIME AT SECOND STAGE, 2Q, q=00

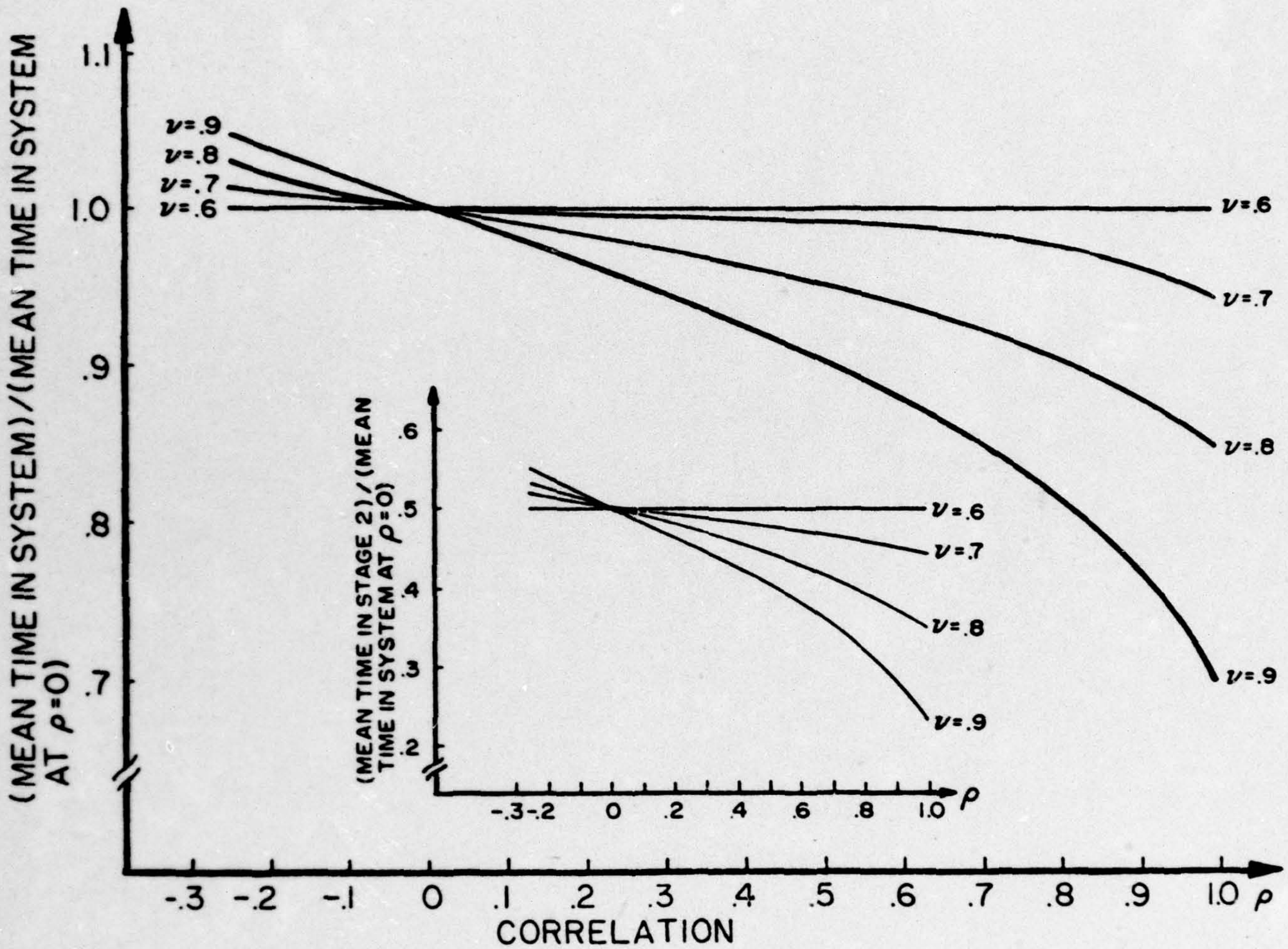


FIGURE 4. RATIO OF MEAN WAITING TIMES, 2Q, q = ∞.

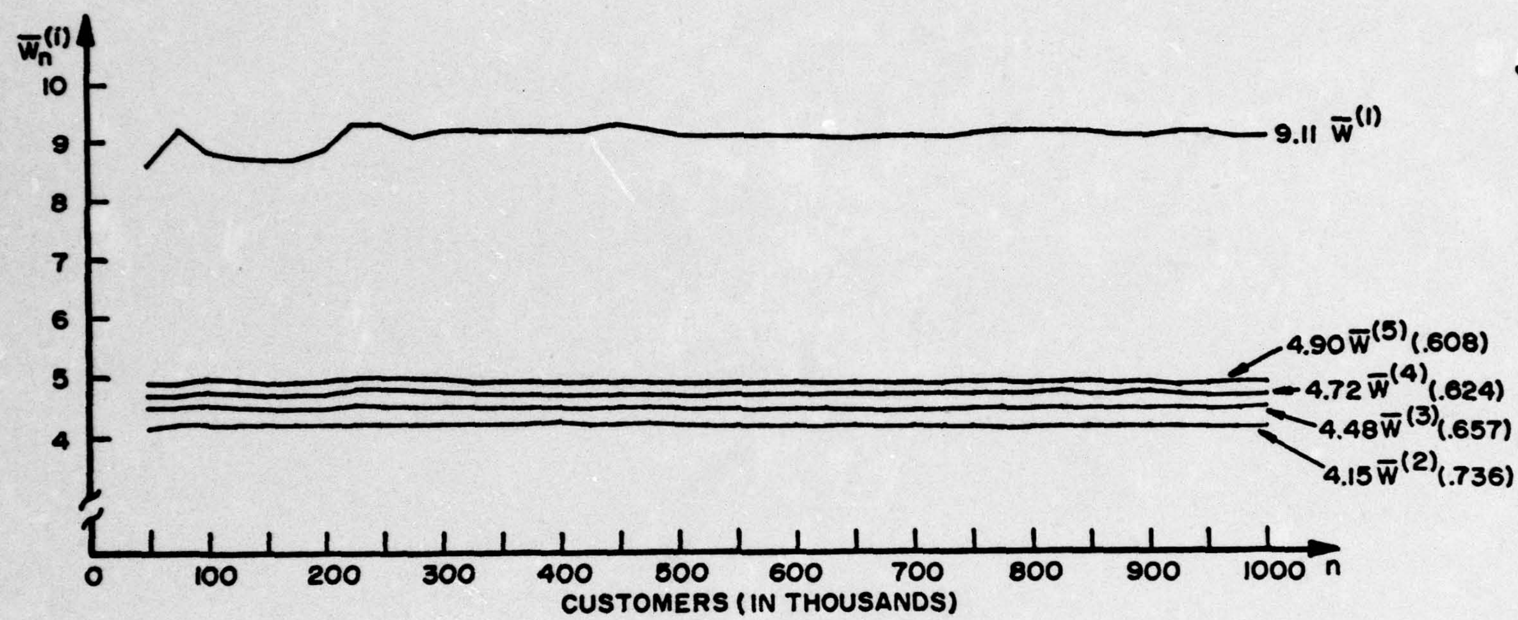


FIGURE 5a. MEAN WAITING TIME, 5Q, $q=\infty$, $\rho=1.0$, $\nu=0.9$.

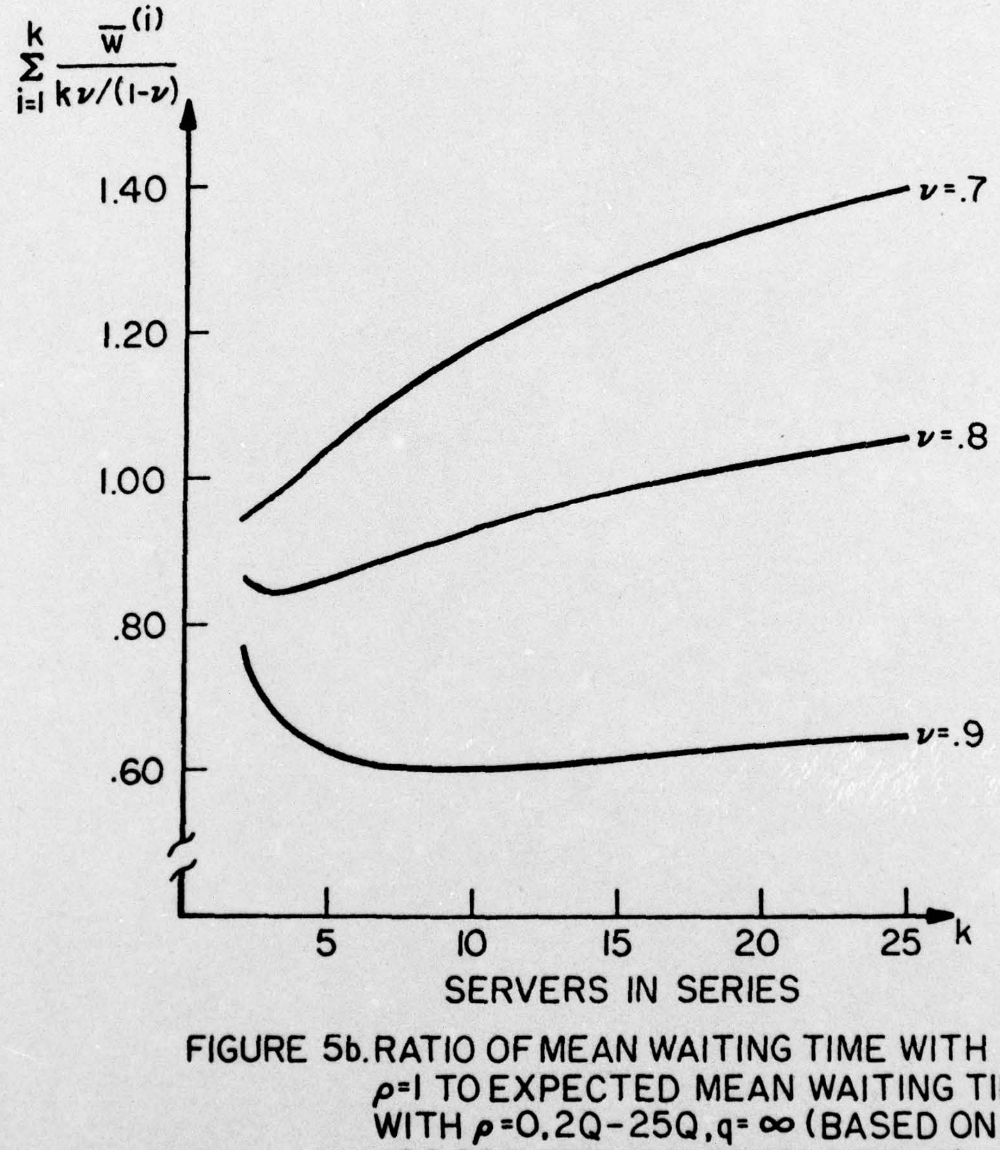


FIGURE 5b. RATIO OF MEAN WAITING TIME WITH $\rho=1$ TO EXPECTED MEAN WAITING TIME WITH $\rho = 0.2Q-25Q$, $q = \infty$ (BASED ON 100,000 SERVICE COMPLETIONS).

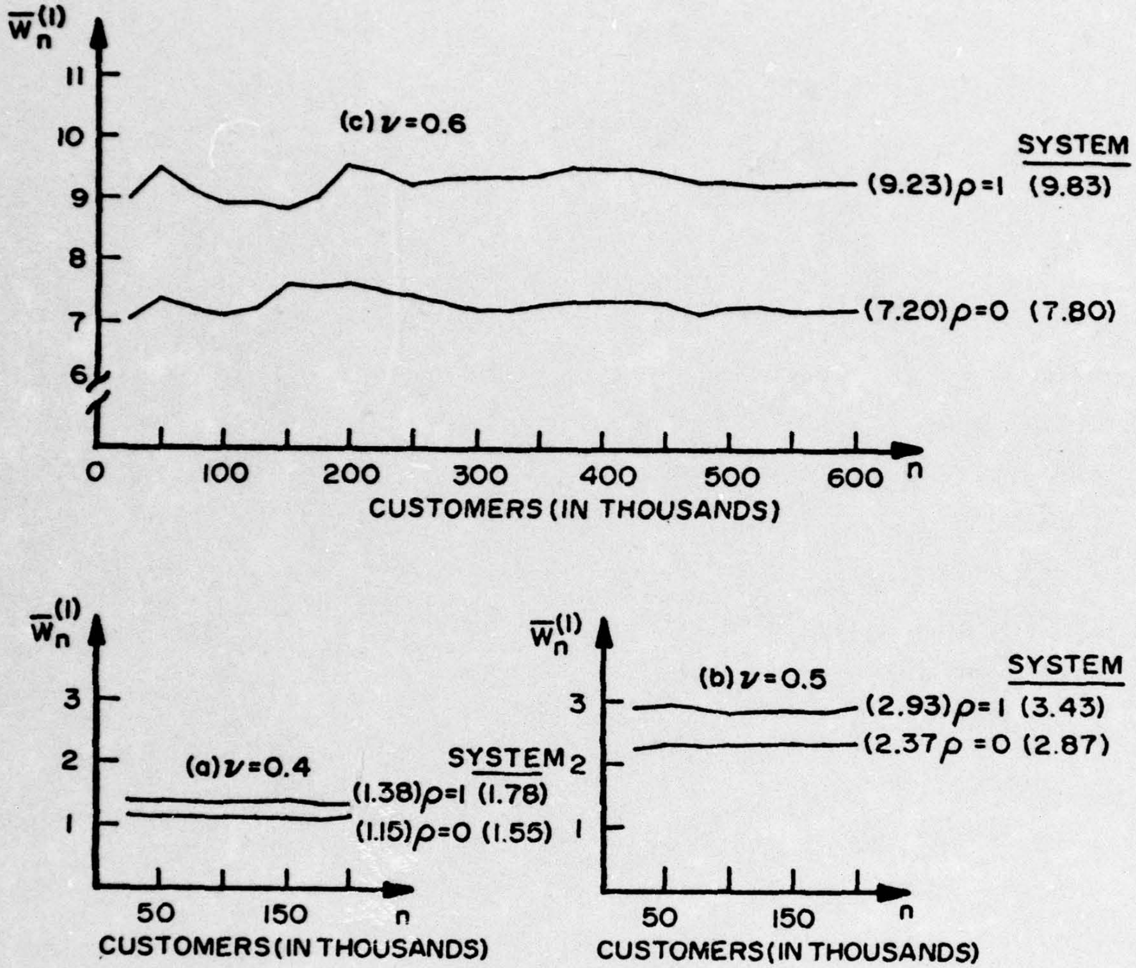


FIGURE 6. MEAN WAITING TIME AT FIRST STAGE, 2Q, q=1.

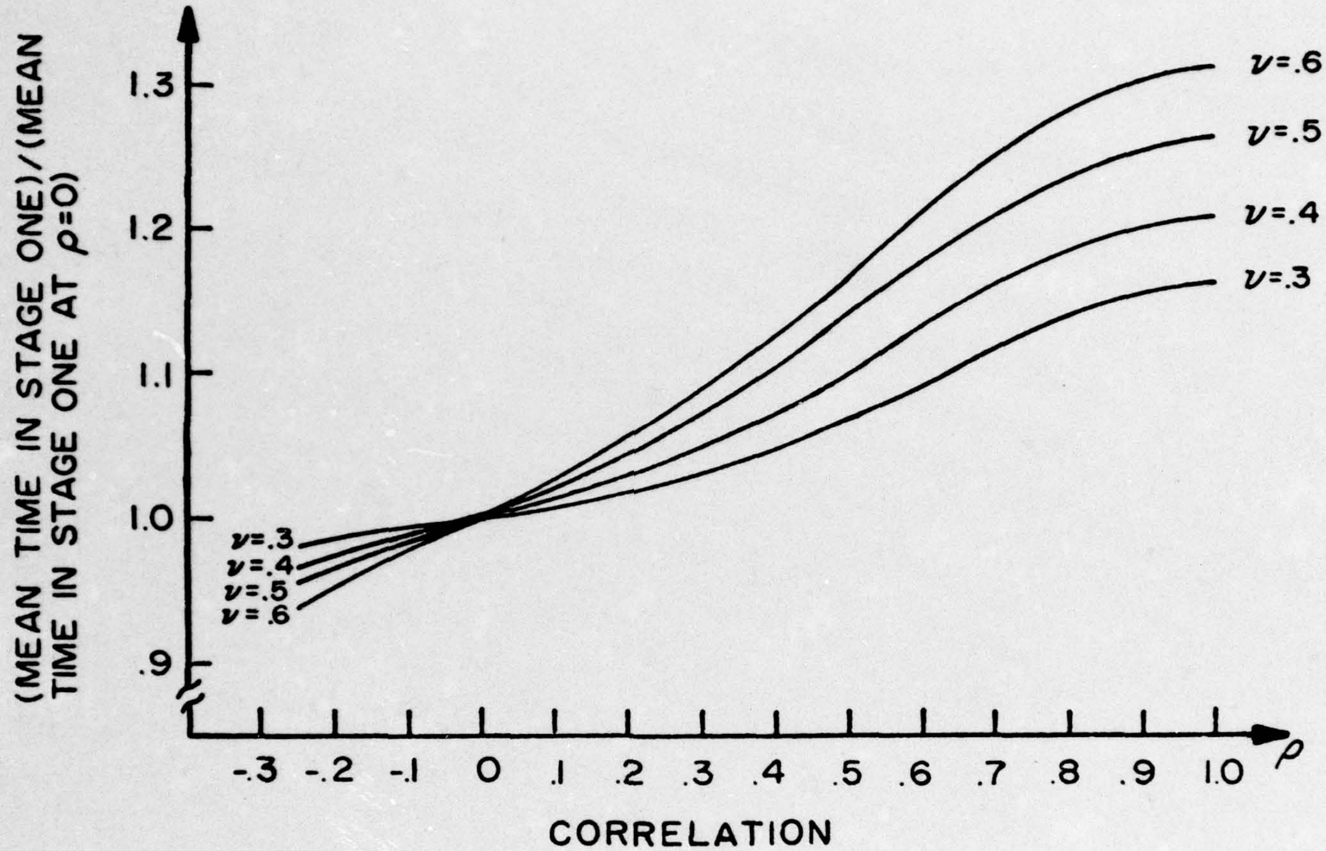


FIGURE 7. RATIO OF MEAN WAITING TIMES AT FIRST STAGE, 2Q, q=1

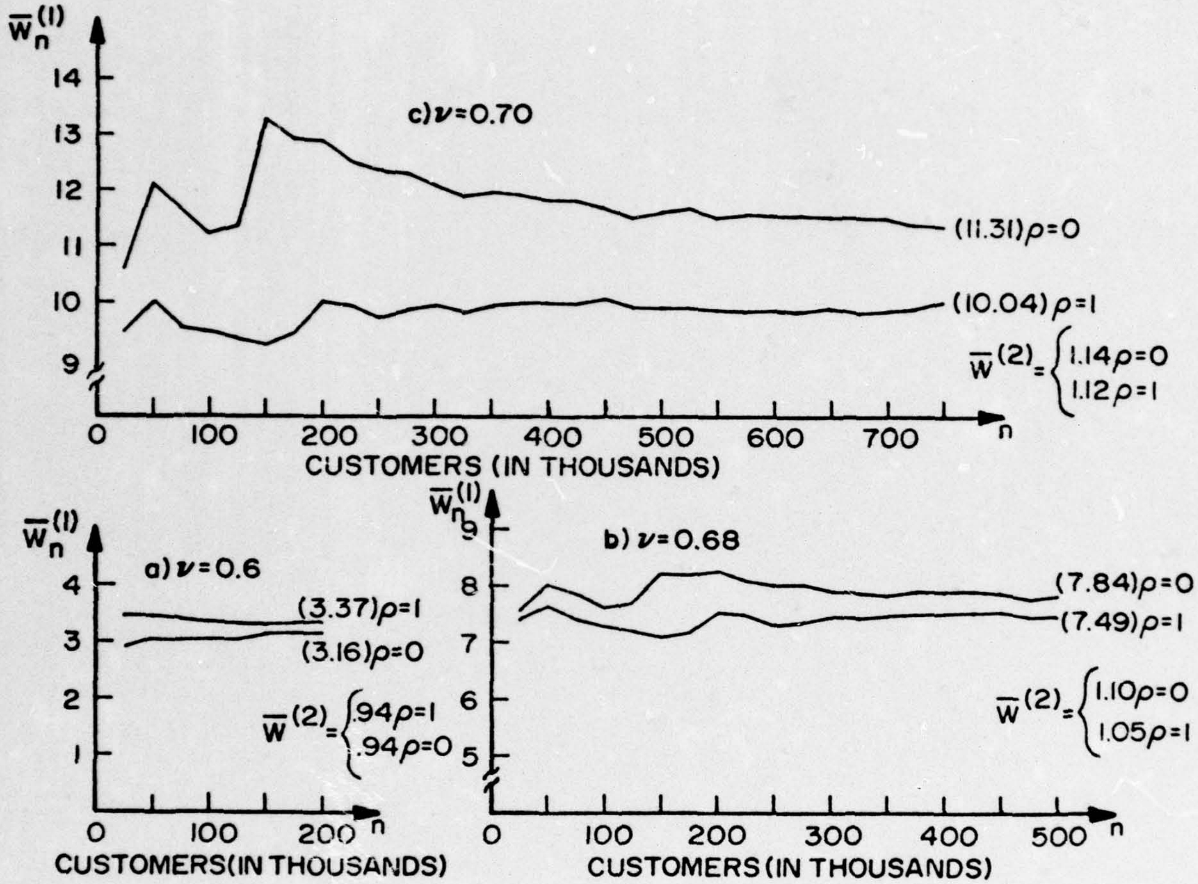


FIGURE 8. MEAN WAITING TIME AT FIRST STAGE, 2Q, q = 2.

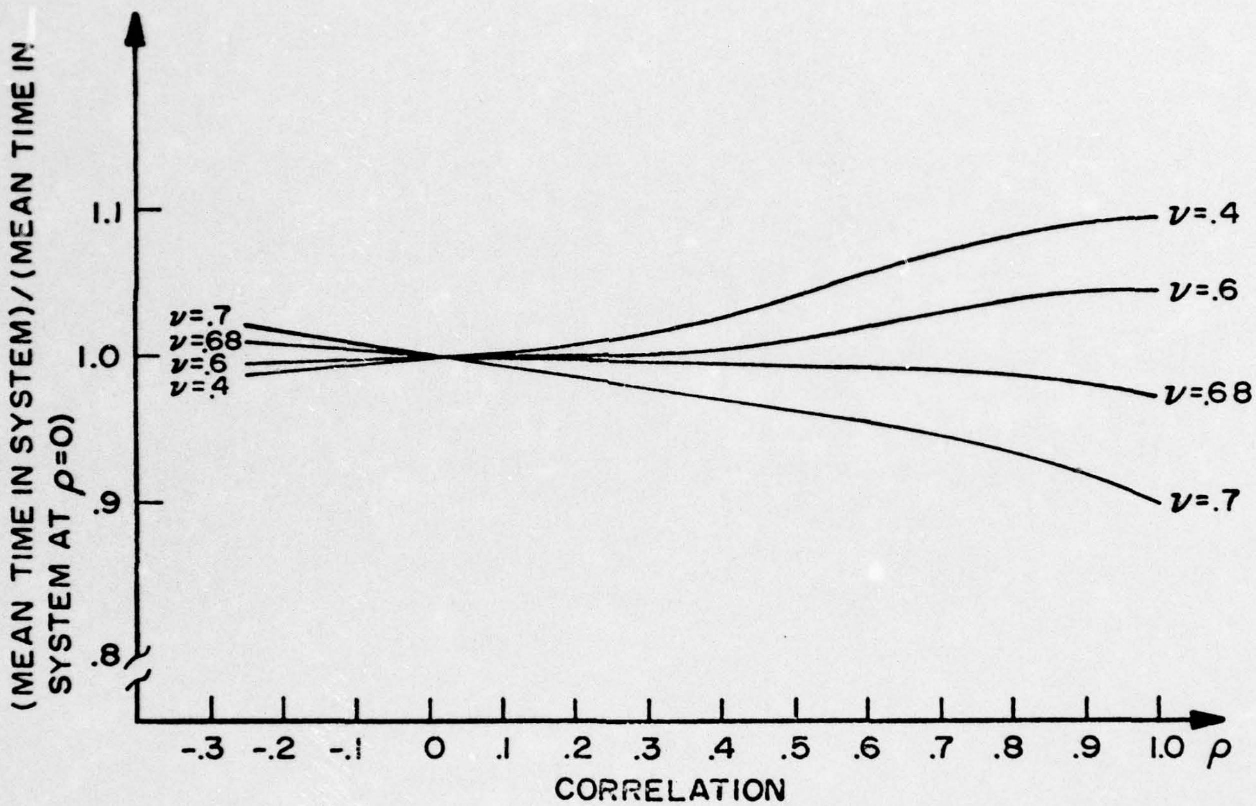


FIGURE 9. RATIO OF MEAN WAITING TIMES IN SYSTEM, 2Q, q= 2.

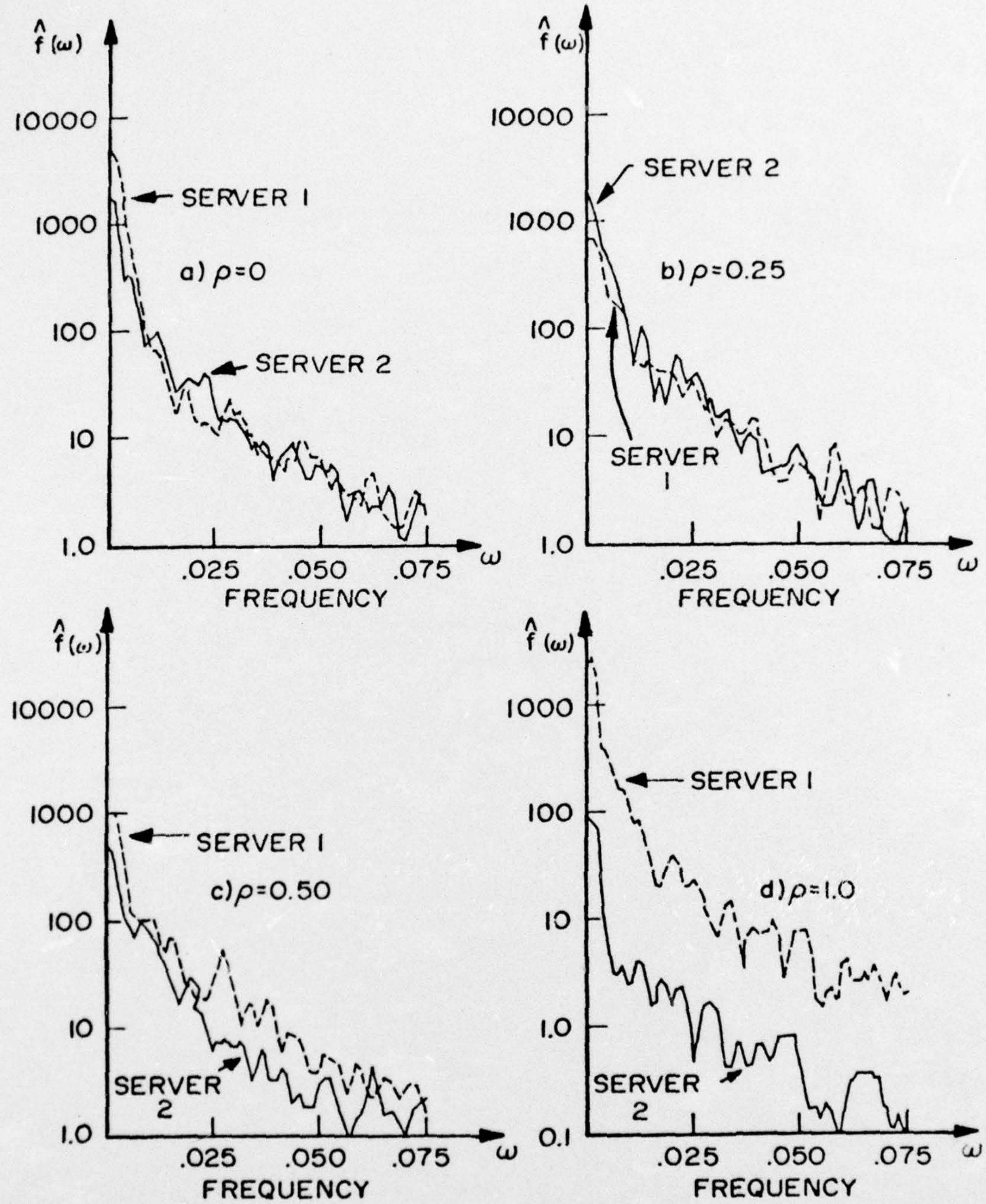


FIGURE 10. A PORTION OF WAITING TIME SPECTRA, 2Q, ν=0.9, q=∞.

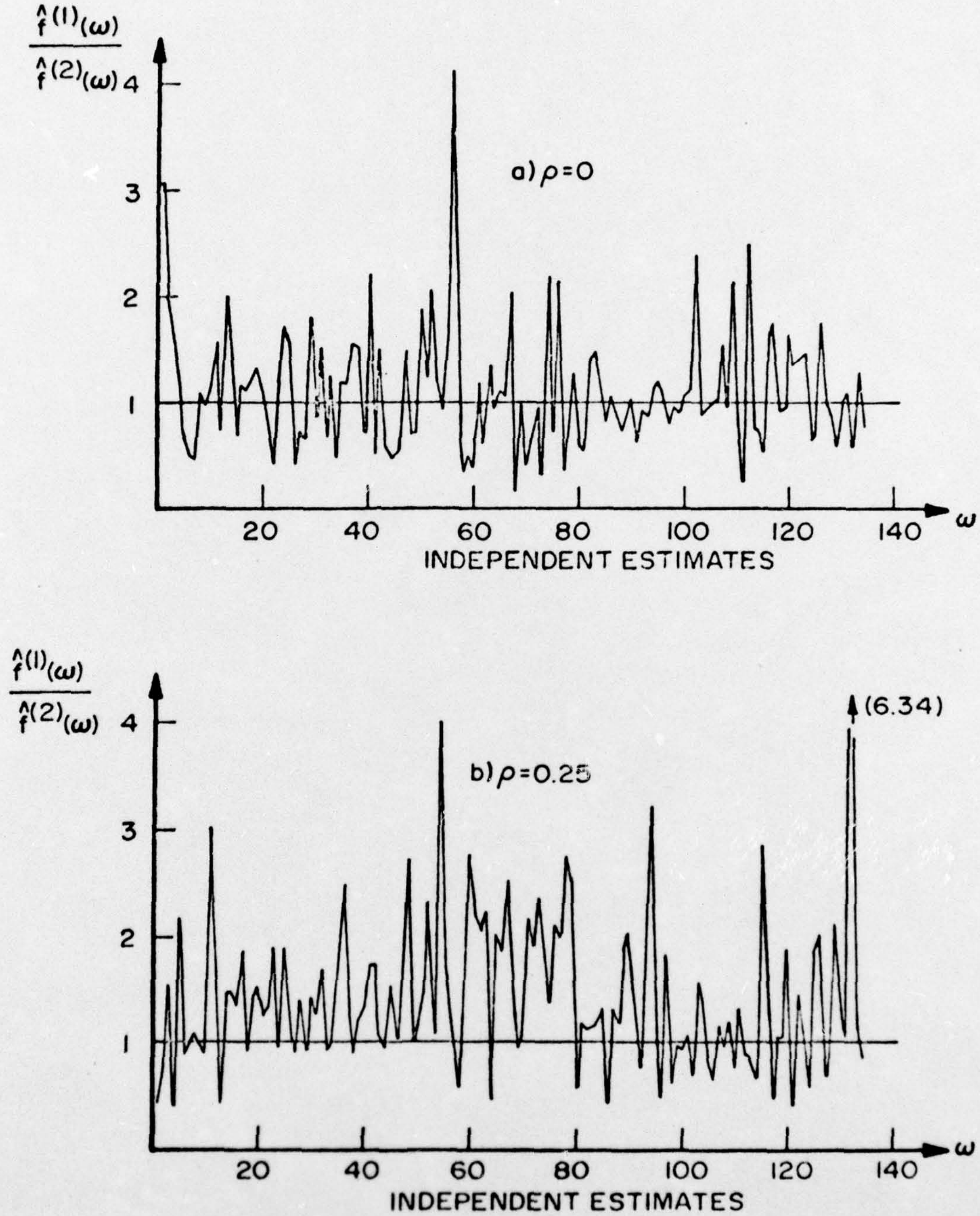


FIGURE 11. RATIO OF SAMPLE POWER SPECTRAL ESTIMATES, 2Q, $\nu=0.9$, $q=\infty$.

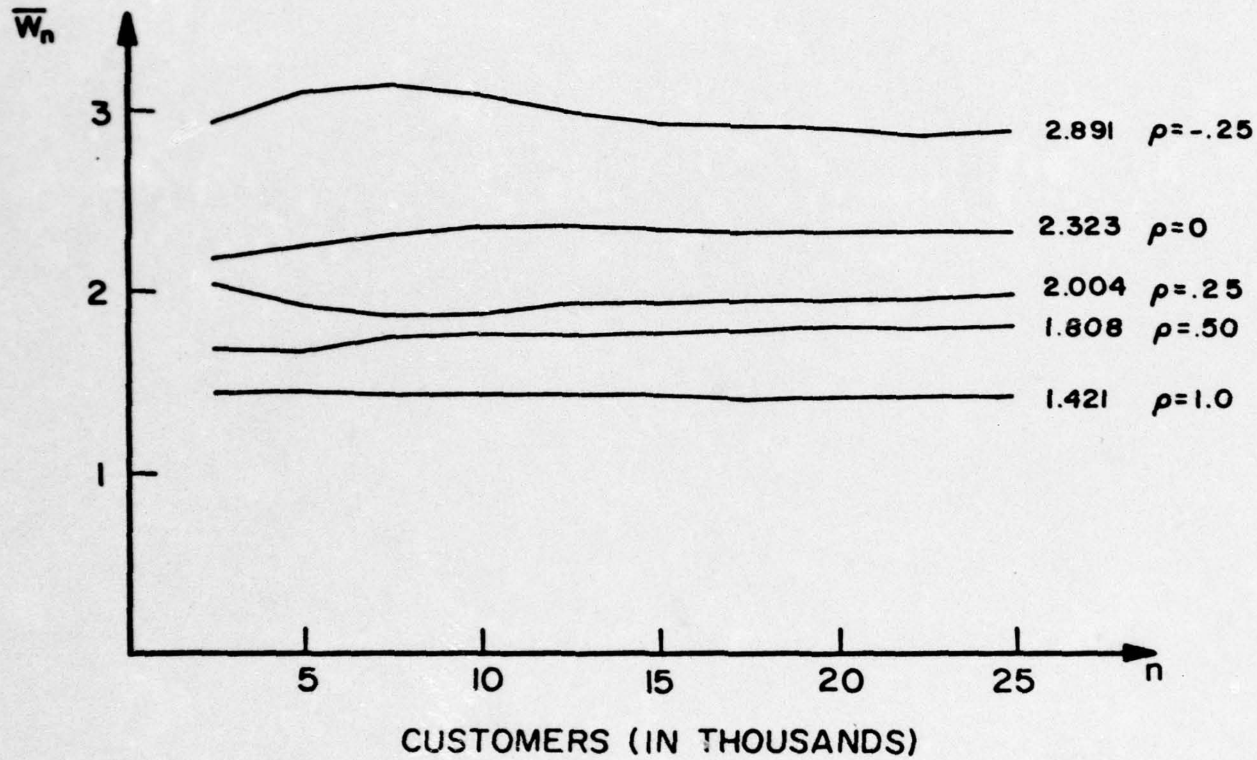


FIGURE 12. MEAN WAITING TIMES, $\nu=0.7$.

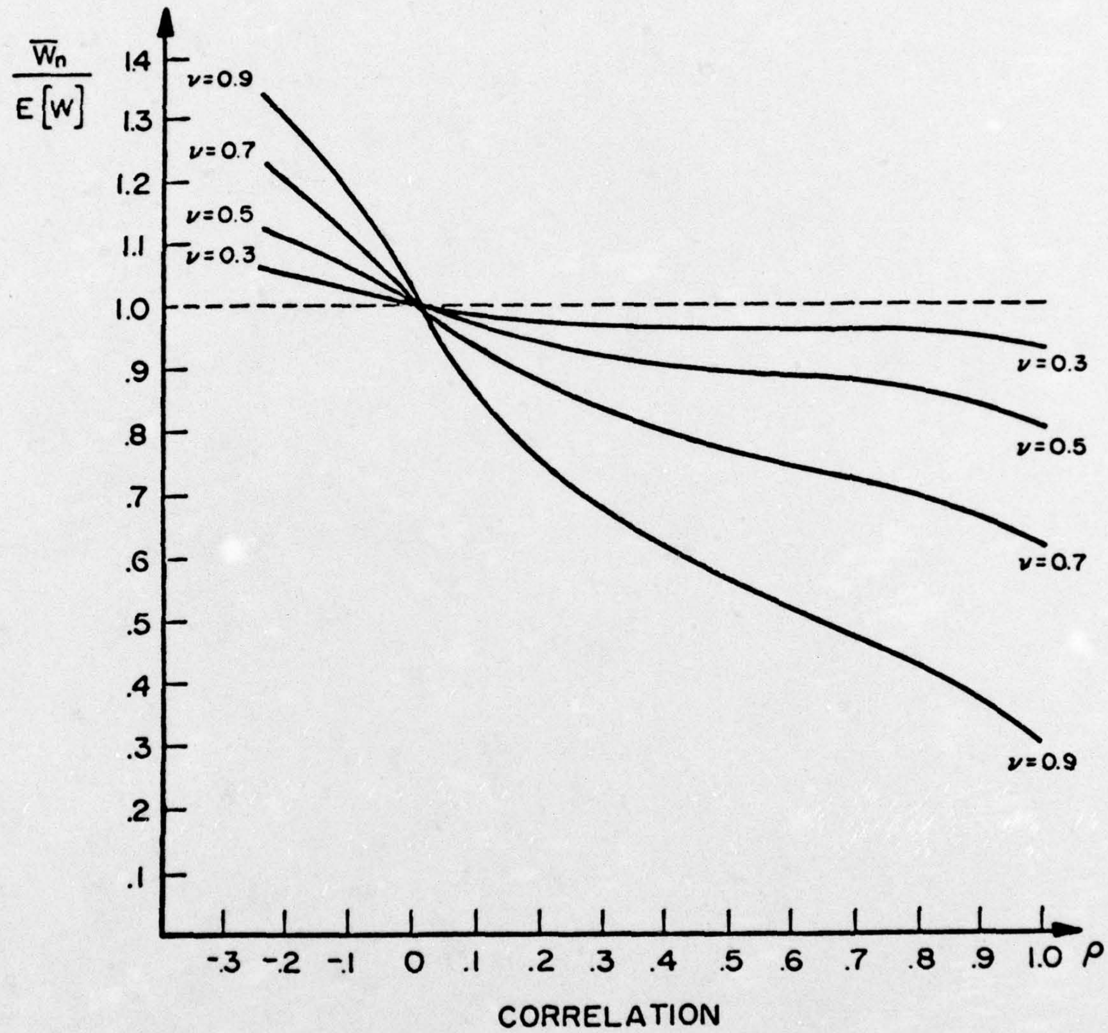


FIGURE 13. RATIO OF MEAN WAITING TIME AT $\rho \neq 0$ TO EXPECTED WAITING TIME AT $\rho = 0$.

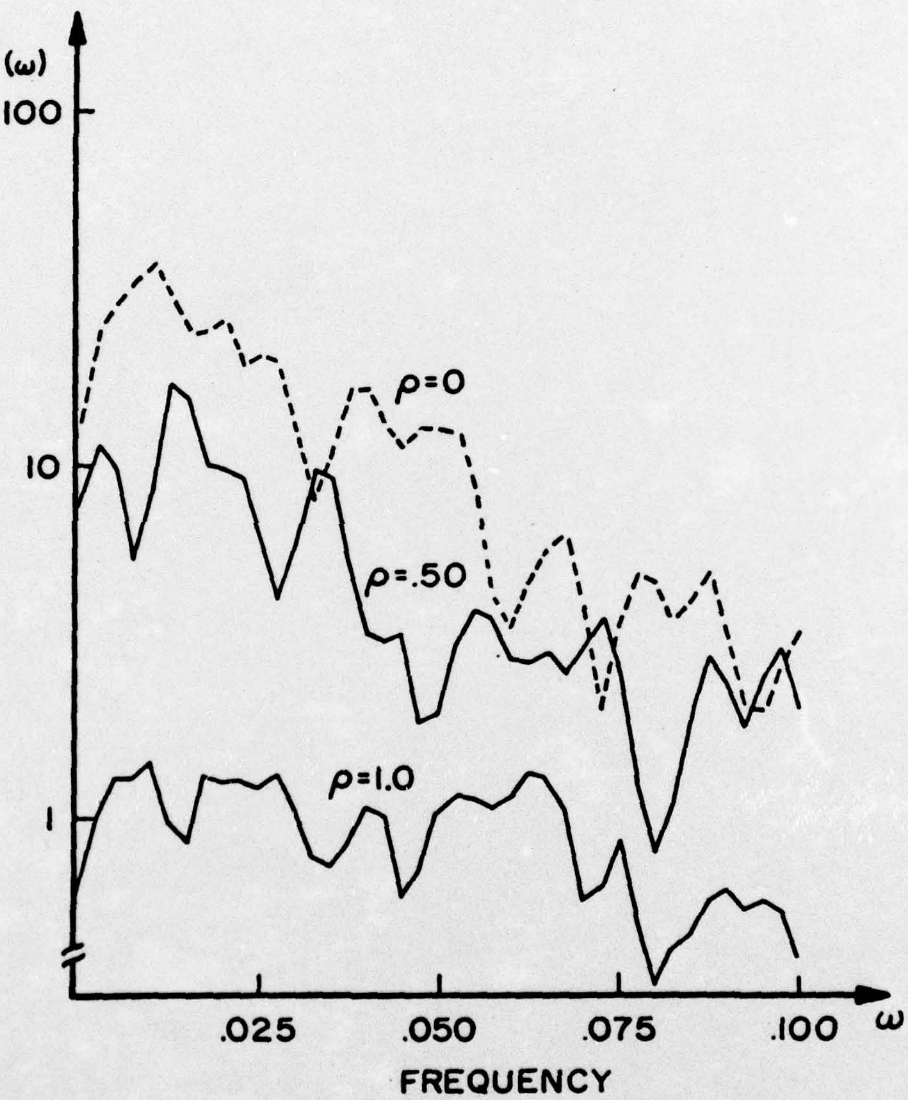


FIGURE 14. A PORTION OF WAITING TIME SPECTRA, $\nu = 0.7$.

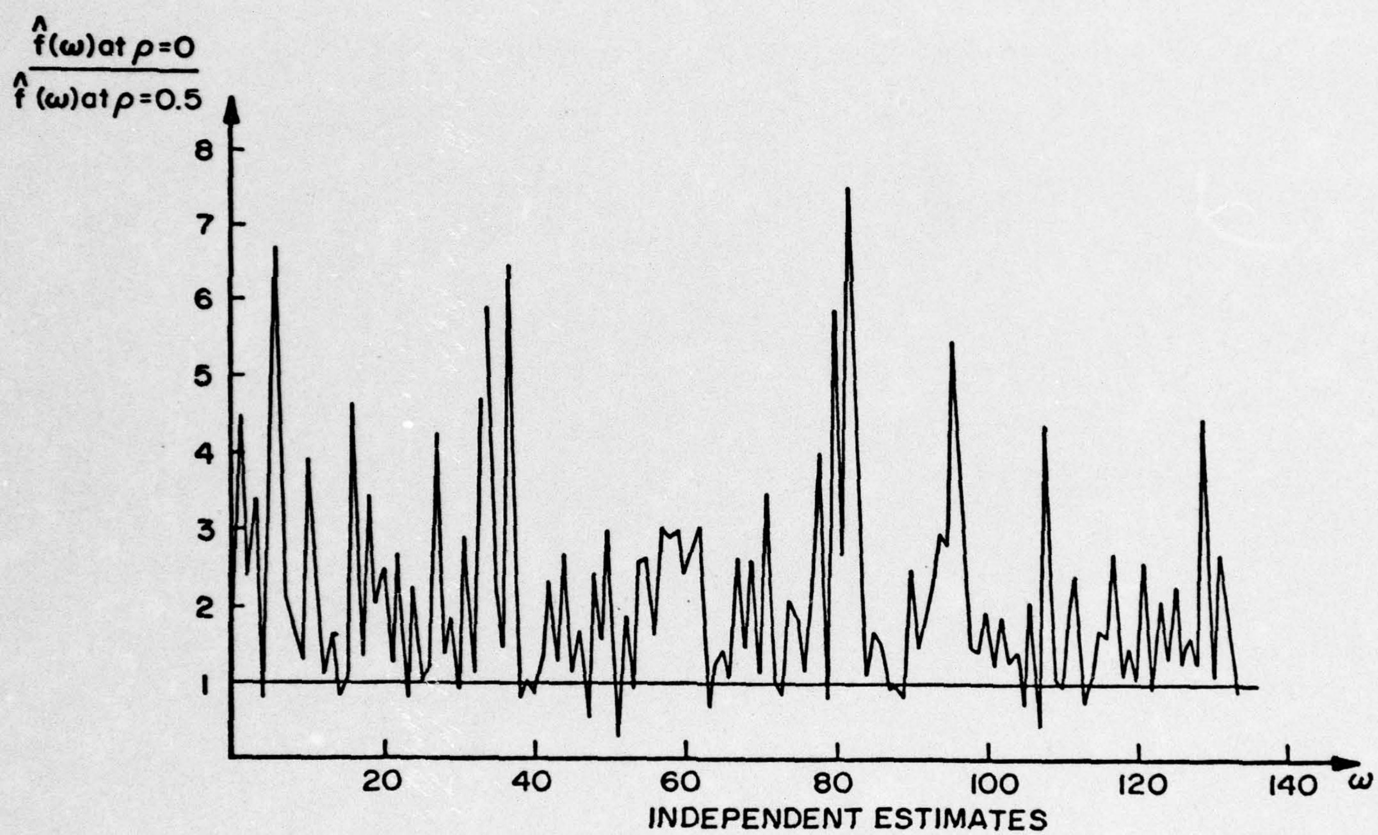


FIGURE 15. RATIO OF SAMPLE POWER SPECTRAL ESTIMATES, ν=0.7

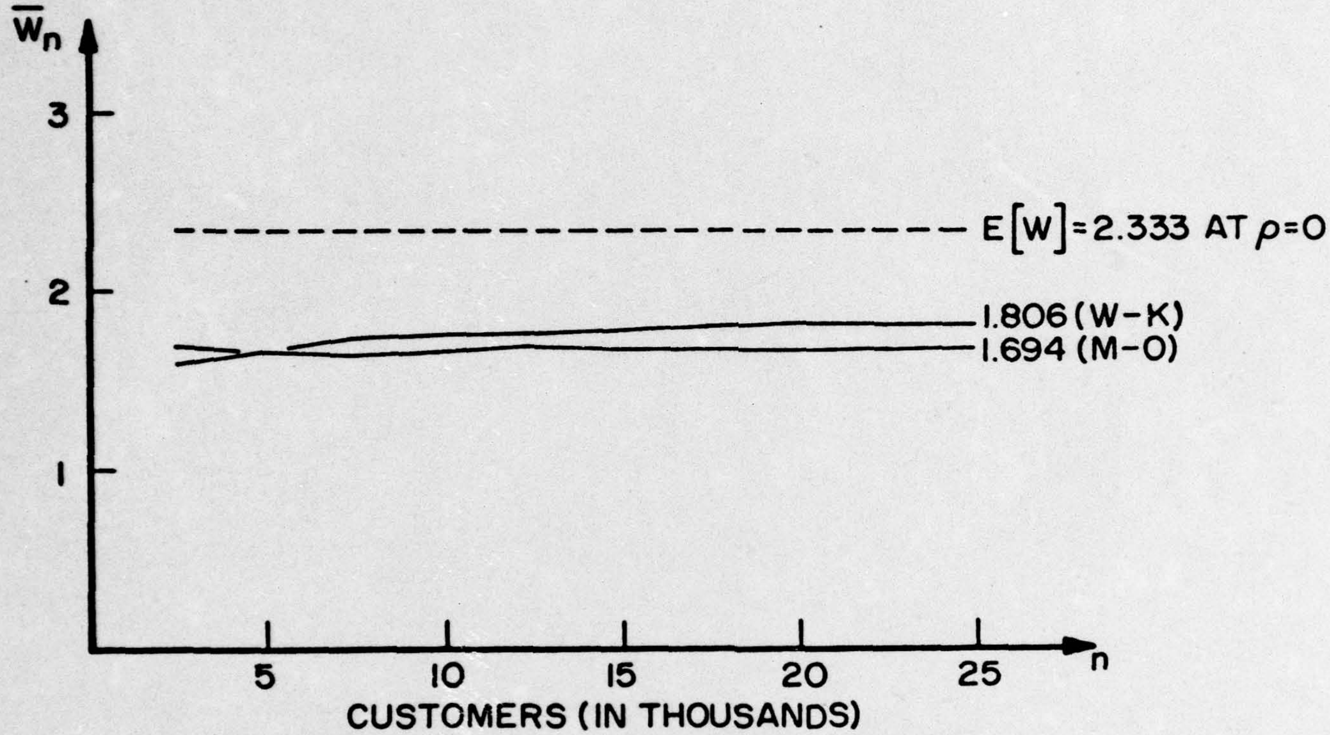


FIGURE 16. COMPARISON OF MEAN WAITING TIMES FOR MARSHALL-OLKIN AND WICKSELL-KIBBLE MODELS, $\rho=0.5, \nu=0.7$